SHADOW PRICES AND WELL-POSEDNESS IN THE PROBLEM OF OPTIMAL INVESTMENT AND CONSUMPTION WITH TRANSACTION COSTS

JIN HYUK CHOI, MIHAI SÎRBU AND GORDAN ŽITKOVIĆ

ABSTRACT. We revisit the optimal investment and consumption model of Davis and Norman (1990) and Shreve and Soner (1994), following a shadow-price approach similar to that of Kallsen and Muhle-Karbe (2010). Making use of the completeness of the model without transaction costs, we reformulate and reduce the Hamilton-Jacobi-Bellman equation for this singular stochastic control problem to a non-standard free-boundary problem for a first-order ODE with an integral constraint. Having shown that the free boundary problem has a smooth solution, we use it to construct the solution of the original optimal investment/consumption problem in a self-contained manner and without any recourse to the dynamic programming principle. Furthermore, we provide an explicit characterization of model parameters for which the value function is finite.

1. INTRODUCTION

Ever since the seminal work of Merton (see [Mer69] and [Mer71]), the problem of dynamic optimal investment and consumption occupied a central role in mathematical finance and financial economics. Merton himself, together with many of the researchers that followed him, made the simplifying assumption of *no market frictions*: there are no transaction costs, borrowing and lending occur at the same interest rate, the assets can be bought and sold immediately in any quantity and at the same price (perfect liquidity), etc. Among those, transaction costs are (arguably) among the most important and (demonstrably) the most studied.

1.1. **Existing work.** The problem of optimal investment where transactions cost are present has received (and continues to receive) considerable attention. Following the early work of Constantinides and Magill [CM76], Davis and Norman [DN90] considered a risky asset driven by the geometric Brownian Motion for which proportional transactions costs are levied on each transaction. These authors formulated the optimal investment/consumption problem as a singular stochastic control problem, and approached it using the method of dynamic programming. Very early in the game it has been intuited, and later proved to varying degrees of rigor, that the optimal strategy has the following general form:

(1) The investor should not trade at all as long as his/her holdings stay within the so-called "no-trade region" - a wedge around the Merton-proportion line.
(2) Outside the "no-trade region' , the investor should trade so as to reach the no-trading region as soon as possible, and, then, adjust the portfolio in a minimal way in order not to leave it ever after.

Such a strategy first appeared in [CM76] and was later made more precise in [DN90]. The analysis of [DN90] was subsequently complemented by that of Shreve and Soner [SS94] who removed various technical conditions, and clarified the key arguments using the technique of viscosity solutions. Still, even in [SS94], technical conditions needed to be imposed. Most notably, the analysis there assumes that *the problem is well posed*, i.e., that the value function is finite; no necessary and

1

sufficient condition for this assumption, in terms of the parameters of the model, is given in [SS94]. In fact, to the best of our knowledge, the present paper provides the first such characterization.

More recently, Kallsen and Muhle Karbe [KMK10] approached the problem using the concept of a *shadow price*, first introduced by [JK95] and [LPS98]. Roughly speaking, the shadow-price approach amounts to comparing the problem *with* transaction costs to a family of similar problems, but *without* transaction costs, whose risky-asset prices lie between the bid and ask prices of the original model. The most unfavorable of these prices is expected to yield the same utility as the original problem where transaction costs are paid. As shown in [KMK10], this approach works quite well for the case of the logarithmic utility, which admits an explicit solution of the problem without transaction costs in a very general class of not-necessarily Markovian models. The fact that the logarithmic utility is the only member of the CRRA (power) family of utility functions with that property makes a direct extension of their techniques seem difficult to implement. Very recently, and in parallel with our work, partial results in this direction have been obtained by Herczegh and Prokaj [HP11] whose approach (and the intuition behind it) is based on the second-order nonlinear free-boundary HJB equation of [SS94], and applies only to a rather restrictive range of parameters.

1.2. **Our contributions.** Our results apply to the model introduced [DN90] or [SS94], and deal with *general power-utility functions* and *general values of the parameters*. It is based on the shadow-price approach, but quite different in philosophy and execution from that of either [KMK10] or [HP11]. Our contributions can be divided into two groups:

Novel treatment and proofs of, as well as insights into the existing results. We provide a new and complete path to the solution to the optimal investment/consumption problem with transaction costs and power-type utilities. Our approach, based on the notion of the shadow price, is fully self-contained, does not rely on the dynamic programming principle and expresses all the features of the solution in terms of a solution to a single, constrained free-boundary problem for a one-dimensional first-order ODE. This way, it is able to distinguish between various parameter regimes which remained hidden under the more abstract approach of [DN90] and [SS94]. Interestingly, most of those regimes turn out to be "singular", in the sense that our first-order ODE develops a singularity in the right-hand side. While we are able to treat them fully, those cases require a much more delicate and insightful analysis. The results of both [KMK10] and [HP11] apply only to the parameter regimes where no singularity is present.

New results. One of the advantages of our approach is that it allows us to give an explicit characterization of the set of model parameters for which the optimal investment and consumption problem with transaction costs is well posed. As already mentioned above, to the best of our knowledge, such a characterization is new, and not present in the literature.

Not only as another application, but also as an integral part of our proof, we furthermore prove that a shadow price exists whenever the problem is well-posed.

Finally, our techniques can be used to provide precise regularity information about all of the analytic ingredients, the value function being one of them. Somewhat surprisingly, we observe that in the singular case these are not always real-analytic, even when considered away from the free boundary.

1.3. **The organization of the paper.** The set-up and the main results are presented in Section 2. In Section 3 we describe the intuition and some technical considerations leading to our non-standard free-boundary problem. In Section 4, we prove a verification-type result, i.e., show how to solve the singular control problem, assuming that a smooth-enough solution for the free-boundary equation is available. The proof of existence of such a smooth solution is the most involved part of the paper. In order to make our presentation easier to parse, we split this proof into two parts. Section 5 presents the main ideas of the proof, accompanied by graphical illustrations. The rigorous proofs follow in Section 6.

2. The Problem and the Main Results

2.1. **The Market.** We consider a model of a financial market in which the price process $\{S_t\}_{t \in [0,\infty)}$ of the risky asset (form simplicity called the "stock") is given by

$$dS_t = S_t(\mu \, dt + \sigma \, dB_t), \ t \in [0,\infty) \text{ with } S_0 > 0.$$

Here, B is a standard Brownian motion, and $\mu > 0$ and $\sigma > 0$ are constants - parameters of the model. The information structure is given by the natural saturated filtration generated by B. An economic agent starts with η_S shares of the stock and η_B units of an interestless bond and invests in the two securities dynamically. Transaction costs are not assumed away, and we model them as proportional to the size of the transaction. More precisely, they are determined by two constants $\underline{\lambda} \in (0,1)$ and $\overline{\lambda} > 0$: one gets only $\underline{S}_t = (1-\underline{\lambda})S_t$ for one share of the stock, but pays $\overline{S}_t = (1+\overline{\lambda})S_t$ for it.

2.2. **Solvency and Admissible Strategies.** We assume that the agent's initial position (η_B, η_S) is **strictly solvent**, which means that it can be liquidated to a positive cash position. More precisely, we assume that $\text{Liq}(\eta_B, \eta_S, \underline{S}_0, \overline{S}_0) > 0$ where

$$\text{Liq}(\varphi^0, \varphi, \underline{s}, \overline{s}) = \varphi^0 + \varphi^+ \underline{s} - \varphi^- \overline{s}. \tag{2.1}$$

The agent's **(consumption/trading) strategy** is described by a triple (φ^0, φ, c) of **optional** processes such that φ and φ^0 are right-continuous and of finite variation and c is nonnegative and locally integrable, a.s. The processes φ^0 and φ have the meaning of the amount of cash held in the money market and the number of shares in the risky asset, respectively, while c is the consumption rate.

In order to incorporate the potential initial jump we distinguish between the initial values $(\varphi^0_{0-}, \varphi_{0-})$ and the values (φ^0_0, φ_0) (after which the processes are right-continuous). This is quite typical for optimal investment/consumption strategies, both in frictional and frictionless markets, when the agent initially holds stocks, in addition to bonds. In this spirit, we always assume that $(\varphi^0_{0-}, \varphi_{0-}) = (\eta_B, \eta_S)$.

A strategy (φ^0, φ, c) is said to be **self-financing** if

$$\varphi^0_t = \varphi^0_{0-} - \int_0^t \overline{S}_u d\varphi^\uparrow_u + \int_0^t \underline{S}_u d\varphi^\downarrow_u - \int_0^t c_u \, du, \tag{2.2}$$

where $\varphi = \varphi_{0-} + \varphi^\uparrow - \varphi^\downarrow$ is the pathwise minimal (Hahn-Jordan) decomposition of φ into a difference of two non-decreasing adapted, right-continuous processes, *with possible jumps at time zero*, as we assume that $\varphi^\uparrow_{0-} = \varphi^\downarrow_{0-} = 0$.

The integrals used in (2.2) above, with respect to the (pathwise Stieltjes) measures $d\varphi^\uparrow$ and $d\varphi^\downarrow$ characterized by $d\varphi^\uparrow((a,b]) = \varphi^\uparrow(b) - \varphi^\uparrow(a)$, and $d\varphi^\downarrow((a,b]) = \varphi^\downarrow(b) - \varphi^\downarrow(a)$, for $0 \le a < b < \infty$ together with $d\varphi^\uparrow(\{0\}) = \varphi^\uparrow(0) - \varphi^\uparrow(0-) = \varphi^\uparrow(0)$, and $d\varphi^\downarrow(\{0\}) = \varphi^\downarrow(0) - \varphi^\downarrow(0-) = \varphi^\downarrow(0)$.

A self-financing strategy (φ^0, φ, c) is called **admissible** if its position is always **solvent**, i.e., if

$$\mathrm{Liq}(\varphi_t^0, \varphi_t, \underline{S}_t, \overline{S}_t) \geq 0, \text{ for all } t, \text{ a.s.} \tag{2.3}$$

The set of all admissible strategies with $\varphi_{0-}^0 = \eta_B$ and $\varphi_{0-} = \eta_S$ is denoted by $\mathcal{A}$, and the set of all c such that $(\varphi^0, \varphi, c) \in \mathcal{A}$ for some φ^0 and φ - the so-called **financeable consumption processes** - is denoted by $\mathcal{C}$.

2.3. **Utility functions.** For $p \in (-\infty, 1)$, we consider the **utility function** $U : [0, \infty) \to [-\infty, \infty)$ of the power (CRRA) type. It is defined for $c \geq 0$ by

$$U(c) = \begin{cases} \frac{1}{p}c^p, & c \neq 0, p \neq 0 \\ \log(c), & c \neq 0, p = 0, \end{cases} \text{ and } U(0) = \begin{cases} 0, & p > 0, \\ -\infty, & p \leq 0 \end{cases}$$

Our task is to analyze the optimal investment and consumption problem, with the value

$$u = \sup_{c \in \mathcal{C}} \mathcal{U}(c), \text{ where } \mathcal{U}(c) = \mathbb{E}\left[\int_0^\infty e^{-\delta t} U(c_t)\, dt\right], \tag{2.4}$$

and $\delta > 0$ stands for the (constant) **impatience rate**. As part of the definition of $\mathcal{U}$, we posit that $\mathcal{U}(c) = -\infty$ unless $\mathbb{E}\left[\int_0^\infty e^{-\delta t}(U(c_t))^-\, dt\right] < \infty$.

2.4. **Consistent price processes.** An Itô-process $\tilde{S}$ is called a **consistent price (process)** if $\underline{S}_t \leq \tilde{S}_t \leq \overline{S}_t$, for all $t \geq 0$, a.s.; the set of all consistent prices is denoted by $\mathcal{S}$. For each consistent price $\tilde{S} \in \mathcal{S}$, and a solvent pair of initial holdings $(\eta_B, \eta_S) \in \mathbb{R}^2$ such that $\mathrm{Liq}(\eta_B, \eta_S, \underline{S}_0, \overline{S}_0) \geq 0$, we define the set $\mathcal{A}(\tilde{S})$ of **(frictionless) admissible** strategies (φ^0, φ, c), as we would in the standard frictionless market where the price-process is given by $\tilde{S}$. More precisely, for (φ^0, φ, c) to belong to $\mathcal{A}(\tilde{S})$ it is necessary and sufficient that the following three conditions hold

(i) φ^0, φ and c are progressively measurable, $c_t \geq 0$, a.s., for all $t \in [0, \infty]$,
(ii) $\varphi_{0-}^0 = \eta_B$ and $\varphi_{0-} = \eta_S$, and
(iii) $V_t = \varphi_t^0 + \varphi_t \tilde{S}_t \geq 0$, for all $t \in [0, \infty)$, a.s,. and

$$V_t = V_{0-} + \int_0^t \varphi_u\, d\tilde{S}_u - \int_0^t c_u\, du, t \geq 0, \text{ a.s.} \tag{2.5}$$

The set of processes $\{c_t\}_{t \in [0,\infty)}$ that appear as the third component of an element of $\mathcal{A}(\tilde{S})$ will be denoted by $\mathcal{C}(\tilde{S})$, i.e.,

$$\mathcal{C}(\tilde{S}) = \{c : \text{there exist } \varphi^0, \varphi, \text{ such that } (\varphi^0, \varphi, c) \in \mathcal{A}(\tilde{S})\}.$$

The elements of $\mathcal{C}(\tilde{S})$ can be interpreted as the consumption processes financeable from the initial holding (η_B, η_S) in the frictionless market modeled by $\tilde{S}$. The intuition that the presence of transaction costs can only reduce the collection of financeable consumption processes can be formalized as in the following easy proposition.

Proposition 2.1. $\mathcal{C} \subseteq \mathcal{C}(\tilde{S})$, for each $\tilde{S} \in \mathcal{S}$.

Proof. For $c \in \mathcal{C}$, let (φ^0, φ) be such that $(\varphi^0, \varphi, c) \in \mathcal{A}$. By the self-financing condition (2.2), the fact that $\underline{S}_t \leq \tilde{S}_t \leq \overline{S}_t$ and integration by parts (simplified by the fact that $\tilde{S}$ is continuous), we have

$$-\int_0^t c_u\, du = \varphi_t^0 - \varphi_{0-}^0 + \int_0^t \overline{S}_u d\varphi_u^\uparrow - \int_0^t \underline{S}_u d\varphi_u^\downarrow \geq \varphi_t^0 - \varphi_{0-}^0 + \int_0^t \tilde{S}_u\, d\varphi_u$$

$$= \varphi_t^0 - \varphi_{0-}^0 + \tilde{S}_t \varphi_t - \tilde{S}_0 \varphi_{0-} - \int_0^t \varphi_u\, d\tilde{S}_u \tag{2.6}$$

Therefore, by the admissibility criterion (2.3), we have

$$\eta_B + \tilde{S}_0 \eta_S + \int_0^t \varphi_u \, d\tilde{S}_u - \int_0^t c_u \, du \geq \varphi_t^0 + \tilde{S}_t \varphi_t \geq 0. \tag{2.7}$$

It remains to set $\tilde{\varphi} = \varphi$ and $\tilde{\varphi}_t^0 = \eta_B + \int_0^t \tilde{\varphi}_u \, d\tilde{S}_u - \int_0^t c_u \, du - \tilde{\varphi}_t \tilde{S}_t + \eta_S \tilde{S}_0$, and observe that (2.7) directly implies (2.5). Thus, $(\tilde{\varphi}^0, \tilde{\varphi}, c) \in \mathcal{A}(\tilde{S})$. $\qquad\square$

It will be important in the sequel to be able to check whether an element of $\mathcal{C}(\tilde{S})$ belongs to $\mathcal{C}$. It happens, essentially, when a strategy that finances it "buys" only when $\tilde{S}_t = \overline{S}_t$ and "sells" only when $\tilde{S}_t = \underline{S}_t$. A precise statement is given in the following proposition.

Proposition 2.2. *Given $\tilde{S} \in \mathcal{S}$, let $c \in \mathcal{C}(\tilde{S})$ be such that there exist processes φ^0 and φ such that*
  (1) $(\varphi^0, \varphi, c) \in \mathcal{A}(\tilde{S})$,
  (2) φ is a right-continuous process of finite variation, and
  (3) the Stieltjes measure on $[0, \infty)$ induced by $\varphi^\uparrow$ is carried by $\{\tilde{S}_t = \overline{S}_t\}$ and that induced by $\varphi^\downarrow$ by $\{\tilde{S}_t = \underline{S}_t\}$.
Then, $c \in \mathcal{C}$.

Proof. Let the triplet $(\varphi^0, \varphi, c) \in \mathcal{A}(\tilde{S})$ satisfy the conditions of the proposition. In particular, we have

$$0 = \eta_B + \eta_S \tilde{S}_0 - \varphi_t^0 - \varphi_t \tilde{S}_t + \int_0^t \varphi_u \, d\tilde{S}_u - \int_0^t c_u \, du.$$

Thanks to condition (3), the integration-by-parts formula and the self-financing property (2.2), it follows that

$$\varphi_t^0 = \eta_B - \int_0^t \tilde{S}_t \, d\varphi_u - \int_0^t c_u \, du = \eta_B - \int_0^t \tilde{S}_t \, d\varphi_u^\uparrow + \int_0^t \tilde{S}_t \, d\varphi_u^\downarrow - \int_0^t c_u \, du$$

$$= \eta_B - \int_0^t \overline{S}_t \, d\varphi_u^\uparrow + \int_0^t \underline{S}_t \, d\varphi_u^\downarrow - \int_0^t c_u \, du.$$

Hence, $c \in \mathcal{C}$. $\qquad\square$

2.5. **Shadow Problems.** For each consistent price process $\tilde{S}$, we define an auxiliary optimal-consumption problem - called the $\tilde{S}$**-problem**, with the value $u(\tilde{S})$, by

$$u(\tilde{S}) = \sup_{c \in \mathcal{C}(\tilde{S})} \mathcal{U}(c), \text{ so that } u \leq \inf_{\tilde{S} \in \mathcal{S}} u(\tilde{S}), \tag{2.8}$$

where u is defined as in (2.4), and the inequality on the right is implied by Proposition 2.1. In words, each consistent price $\tilde{S}$ affords at least as good an investment opportunity as the original frictional market.

It is in the heart of our approach to show that the duality gap, in fact, closes, i.e., that the inequality in (2.8) becomes an equality; the worst-case shadow problem performs no better than frictional one.

Definition 2.3. A consistent price $\tilde{S}$ is called a **shadow price** if $u = u(\tilde{S})$.

The central idea of the present paper is to look for a shadow price as the minimizer of the right-hand side of (2.8) viewed as a stochastic control problem. More precisely, we turn our attention to a search for an optimizer in the **shadow problem**:

$$\tilde{u} = \inf_{\tilde{S} \in \mathcal{S}} u(\tilde{S}). \tag{2.9}$$

We start by tackling the shadow problem in a formal manner and deriving an analytic object (a free-boundary problem) related to its solution. Next, we show that this free-boundary problem indeed admits a solution and use it to construct the candidate shadow price. Finally, instead of showing that our candidate is indeed an optimizer for (2.9) and that $u = \tilde{u}$, we use the following direct consequence of Proposition 2.2.

Proposition 2.4. *Suppose that for $\tilde{S} \in \mathcal{S}$ there exists a triplet (φ^0, φ, c) such that*

(1) (φ^0, φ, c) satisfies conditions (1), (2) and (3) of Proposition 2.2, and
(2) $\mathcal{U}(c') \leq \mathcal{U}(c)$, for all $c' \in \mathcal{C}(\tilde{S})$,

Then, $\tilde{S}$ is a shadow price.

Remark 2.5. The route we take towards the existence of a shadow price may appear to be somewhat circuitous. It is chosen so as to maximize the intuitive appeal of the method and minimize (already formidable) technical difficulties.

While the remainder of the paper is devoted to the implementation of the above idea, we anticipate its final results here, for the convenience of the reader. An important by-product of our analysis is the *explicit characterization* of those parameter values which result in a well-posed problem (the value function is finite). To the best of our knowledge, such a characterization is not present in the literature, and the finiteness of the value function is either assumed (as in [SS94]) or deduced from rather strong conditions (as in [KMK10]).

Theorem 2.6. *Given the environment parameters $\mu, \sigma \in (0, \infty)$ and the transaction costs $\underline{\lambda} \in (0, 1)$, $\overline{\lambda} > 0$, the following statements are equivalent:*

(1) The problem is well posed, i.e
$$-\infty < u < \infty,$$
whenever $\mathrm{Liq}(\eta_B, \eta_S, \underline{S}_0, \overline{S}_0) > 0$.

(2) The parameters of the model satisfy one of the following three conditions:
- *$p \leq 0$,*
- *$0 < p < 1$ and $\mu < \sqrt{\frac{2\delta(1-p)\sigma^2}{p}}$,*
- *$0 < p < 1$, $\sqrt{\frac{2\delta(1-p)\sigma^2}{p}} \leq \mu < \frac{\delta}{p} + \frac{(1-p)\sigma^2}{2}$ and*

$$C(\mu, \sigma, p, \delta) < \log(\tfrac{1+\overline{\lambda}}{1-\underline{\lambda}}),$$

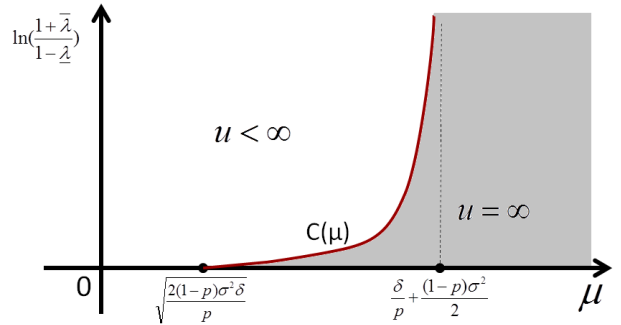*where the function $C(\cdot, \cdot, \cdot, \cdot)$ is given by (6.7) in a closed form.*

FIGURE 1. The well-posedness region.

Remark 2.7. For $\sigma = 0$, the third condition in (2) above reduces to a well-known condition of Shreve and Soner. Indeed, the entire Section 12 in [SS94], culminating in Theorem 12.2, p. 677, is devoted to the well-posedness problem with two bonds (i.e, with $\sigma = 0$).

As demonstrated by our second main result, the shadow-price approach not only allows us to fully characterize the conditions under which a solution to the frictional optimal investment/consumption problem exists, but it also sheds light on its form and regularity.

Theorem 2.8. *Given the parameters $\mu, \sigma \in (0, \infty)$ and the transaction costs $\underline{\lambda} \in (0, 1)$, $\overline{\lambda} > 0$, we assume that well-posedness conditions of Theorem 2.6 hold. Then*

(1) *There exist constants $\underline{x}, \overline{x}$ with $0 < \underline{x} < \overline{x}$ and a function $g \in C^2[\underline{x}, \overline{x}]$ such that*

 (a) *$g'(x) > 0$ for $x \in (\underline{x}, \overline{x})$, and g satisfies the equation*

$$\inf_{\Sigma, \theta \in \mathbb{R}} \left(\tfrac{1}{2}\Sigma^2 \tfrac{x}{g'(x)} - \alpha_q(\Sigma, \theta)x - \beta(\theta)g(x) + \gamma(\theta) \right) = 0, \ x \in (\underline{x}, \overline{x}), \tag{2.10}$$

 where

$$q = \tfrac{p}{1-p}, \ \alpha_q(\Sigma, \theta) = \theta\sigma - \mu - \Sigma\left(\tfrac{1}{2}\Sigma + \sigma - \theta(1+q)\right),$$

$$\beta(\theta) = (1+q)\left(\delta - \tfrac{1}{2}q\theta^2\right), \ and \ \gamma(\theta) = \begin{cases} \tfrac{1}{2}\theta^2, & p = 0, \\ \operatorname{sgn}(p), & p \neq 0. \end{cases} \tag{2.11}$$

 (b) *the following boundary/integral conditions are satisfied:*

$$g'(\underline{x}+) = g'(\overline{x}-) = 0 \ and \ \int_{\underline{x}}^{\overline{x}} \tfrac{g'(x)}{x}\,dx = \log\left(\tfrac{1+\overline{\lambda}}{1-\underline{\lambda}}\right). \tag{2.12}$$

 (c) *The function $h : [\underline{x}, \overline{x}] \to \mathbb{R}$, defined by*

$$h(x) = \begin{cases} (1-x)g'(x) + 1, & p = 0, \\ qg(x)\left(g'(x) + 1\right) - (q+1)xg'(x), & p \neq 0, \end{cases} \tag{2.13}$$

 admits no zeros on $[\underline{x}, \overline{x}]$.

(2) *For any (η_B, η_S) such that*

$$\operatorname{Liq}(\eta_B, \eta_S, \underline{S}_0, \overline{S}_0) > 0, \tag{2.14}$$

 there exists a shadow price $\{\tilde{S}_t\}_{t \in [0, \infty)}$, of the form $\tilde{S}_t = S_t e^{f(X_t^{\hat{x}})}$, where

 - *$f(x) = \underline{y} + \int_x^{\overline{x}} \tfrac{g'(t)}{t}\,dt$, for $x \in [\underline{x}, \overline{x}]$,*
 - *the value of the constant $\hat{x}$ is determined as in Proposition 4.5, and*
 - *$X^{\hat{x}}$ is the unique solution of reflected SDE (4.3) with $X_0^{\hat{x}} = \hat{x}$.*

(3) *For any (η_B, η_S) satisfying (2.14), the value u and an optimal investment/consumption strategy $(\hat{\varphi}^0, \hat{\varphi}, \hat{c})$ for the main problem (2.4) are given by*

$$u = \hat{u}(\eta_B, \eta_S; \hat{x}), \quad (\hat{\varphi}_t^0, \hat{\varphi}_t, \hat{c}_t) = (\hat{\varphi}_t^{0, \hat{x}}, \hat{\varphi}_t^{\hat{x}}, \hat{c}_t^{\hat{x}}),$$

 where $\hat{x}$ is defined in Proposition 4.5, and $\hat{u}, \hat{\varphi}_t^{0, x}, \hat{\varphi}_t^x$ and $\hat{c}_t^x$ in Lemma 4.4.

Remark 2.9. In (2.12), if $(\underline{x}, g(\underline{x})) = P$ (P is a singular point described in Section 5), then the condition $g'(\underline{x}) = 0$ can be violated. For this exceptional case, Proposition 4.2 is still valid: More precisely, in the part (2) of Proposition 4.2, we need to show that $\frac{g'(X_t^x)}{X_t^x}d\Phi_t^x \equiv 0$. If $(\underline{x}, g(\underline{x})) = P$, in (4.3), the drift is positive and the volatility is zero, thus, we conclude that $d\Phi_t^{x\uparrow} \equiv 0$.

3. A heuristic derivation of a free-boundary problem

The purpose of the present section is to provide a heuristic derivation of a free-boundary problem for a one-dimensional first-order ODE which will later be used to construct a shadow process and the solution of our main problem. With the fully rigorous verification coming later, we often do not pay attention to integrability or measurability conditions and formally push through many steps in this section.

We start by splitting the shadow problem (2.9) according to the starting value of the process $\tilde{S}$:

$$\tilde{u} = \inf_{s_0 \in [(1-\underline{\lambda})S_0, (1+\overline{\lambda})S_0]} \quad \inf_{\tilde{S} \in \mathcal{S}, \tilde{S}_0 = s_0} \quad \sup_{c \in \mathcal{C}(\tilde{S})} \mathcal{U}(c). \tag{3.1}$$

One can significantly simplify the analysis of the above problem by noting that, since each $\tilde{S}$ is a strictly positive Itô-process, we can always choose processes $\Sigma = \Sigma(\tilde{S})$ and $\theta = \theta(\tilde{S})$ such that

$$d\tilde{S}_t = \tilde{S}_t(\sigma + \Sigma_t)\,(dB_t + \theta_t\,dt), \ \ \tilde{S}_0 = s_0.$$

It pays to pass to the logarithmic scale, and introduce the process $Y_t = \log(\tilde{S}_t/S_t)$, whose dynamics is given by

$$dY_t = \alpha_0(\theta_t, \Sigma_t)\,dt + \Sigma_t\,dB_t, \tag{3.2}$$

on the natural domain $Y_t \in [\underline{y}, \overline{y}]$. Here, $\underline{y} = \log(1-\underline{\lambda})$, $\overline{y} = \log(1+\overline{\lambda})$ and the function α_0 is given by (2.11) for $q = 0$. This way, the family of consistent price processes is parametrized by the set

$$\mathcal{P} = \{(y, \Sigma, \theta) \,:\, y \in [\underline{y}, \overline{y}], (\Sigma, \theta) \in \mathcal{P}(y)\},$$

where $\mathcal{P}(y)$ is the set of all pairs of regular-enough processes (Σ, θ) such that the process $\{Y_t\}_{t \in [0,\infty)}$, given by (3.2) and starting at $Y_0 = y$, stays in the interval $[\underline{y}, \overline{y}]$, a.s.

We note that the market modeled by $\tilde{S}$ is complete, and that, thanks to the absence of friction, the agent with the initial holdings (η_B, η_S) will achieve the same utility as the one who immediately liquidates the position, i.e., the one with the initial wealth of $\eta_B + S_0 e^y \eta_S$. Therefore, the standard duality theory suggests that

$$\sup_{c \in \mathcal{C}(\tilde{S})} \mathcal{U}(c) = \inf_{z > 0} \left((\eta_B + S_0 e^y \eta_S)z + \mathcal{V}(z\mathcal{E}(-\theta \cdot B)) \right), \tag{3.3}$$

where $\tilde{S}$ and (y, Σ, θ) are related as above and

$$V(z) = \sup_{c > 0}(U(c) - cz), \quad \mathcal{V}(Z) = \mathbb{E}\left[\int_0^\infty e^{-\delta t} V(e^{\delta t} Z_t)\,dt \right]. \tag{3.4}$$

Remark 3.1. The Legendre-Fenchel transform V of U admits an explicit and simple expression in the case of a power utility. Indeed, we have

$$V(z) = \begin{cases} \frac{1}{q}z^{-q}, & p \neq 0, \\ -1 - \log(z), & p = 0, \end{cases}$$

where $q = p/(1-p)$. The parameter q is the negative of the conjugate exponent of p, i.e., $\frac{1}{p} - \frac{1}{q} = 1$ ($q = 0$, for $p = 0$) and this relationship will be assumed to hold throughout the paper without explicit mention.

Consequently, if we combine (3.1) and (3.3), we obtain the following equality:

$$\tilde{u} = \inf_{(y,z) \in [\underline{y}, \overline{y}] \times (0,\infty)} \left((\eta_B + S_0 e^y \eta_S)z + \inf_{(\Sigma, \theta) \in \mathcal{P}(y)} \mathcal{V}(z\mathcal{E}(-\theta \cdot B)) \right). \tag{3.5}$$

The expression above is particularly convenient because it separates the shadow problem into a stochastic control problem over $\mathcal{P}(y)$, and a (finite-dimensional) optimization problem over y and z, which can be solved separately.

3.1. **A dimensional reduction.** Thanks to homogeneity (log-homogeneity for $p = 0$) of the map $z \mapsto \mathcal{V}(zZ)$, a dimensional reduction is possible in the inner control problem in (3.5). Indeed, with $\hat{\delta} = \delta(1 + q)$, we have

$$\mathcal{V}(z\mathcal{E}(-\theta \cdot B)) = \begin{cases} -\frac{2 + \log(z)}{\delta} + \mathbb{E}\left[\int_0^\infty e^{-\delta t}\left(-\log(\mathcal{E}(-\theta \cdot B)_t)\right) dt\right], & p = 0, \\ \frac{z^{-q}}{q}\, \mathbb{E}\left[\int_0^\infty e^{-\hat{\delta}t}\mathcal{E}(-\theta \cdot B)_t^{-q}\, dt\right], & p \neq 0. \end{cases}$$

Hence,

$$\tilde{u} = \inf_{y \in [\underline{y}, \overline{y}]} \begin{cases} \frac{1}{\delta}\left(-1 + \log\left(\delta(\eta_B + S_0 e^y \eta_S)\right) + w(y)\right), & p = 0 \\ \frac{(\eta_B + S_0 e^y \eta_S)^p}{p}\, |w(y)|^{1-p}, & p \neq 0, \end{cases} \tag{3.6}$$

where

$$w(y) = \inf_{(\Sigma, \theta) \in \mathcal{P}(y)} \begin{cases} \mathbb{E}\left[\int_0^\infty \delta e^{-\delta t}\left(-\log(\mathcal{E}(-\theta \cdot B)_t)\right) dt\right], & p = 0, \\ \mathrm{sgn}(p)\mathbb{E}\left[\int_0^\infty e^{-\hat{\delta}t}\mathcal{E}(-\theta \cdot B)_t^{-q}\, dt\right], & p \neq 0. \end{cases}$$

In the heuristic spirit of the present section, it will be assumed that the processes of the form $\theta \cdot B$ and $\mathcal{E}(q\theta \cdot B)$ are (true) martingales so that the definition of the stochastic exponential and the simple identity

$$\mathcal{E}(-\theta \cdot B)^{-q} = \mathcal{E}(q\theta \cdot B) \exp\left(\tfrac{1}{2}q(1 + q)\int_0^\cdot \theta_u^2\, du\right) \tag{3.7}$$

can be used to simplify the expression for w even further:

$$w(y) = \inf_{(\Sigma, \theta) \in \mathcal{P}(y)} \begin{cases} \frac{1}{2}\mathbb{E}\left[\int_0^\infty e^{-\delta t}\theta_t^2\, dt\right], & p = 0, \\ \mathrm{sgn}(p)\mathbb{E}^{\bar{\mathbb{P}}}\left[\int_0^\infty e^{-\hat{\delta}t}e^{\frac{1}{2}q(1+q)\int_0^t \theta_u^2\, du}\, dt\right], & p \neq 0. \end{cases} \tag{3.8}$$

Here, the measure[1] $\bar{\mathbb{P}}$ is (locally) given by $d\bar{\mathbb{P}} = \mathcal{E}(q\theta \cdot B)\, d\mathbb{P}$. By Girsanov's theorem the process

$$\bar{B} = B - \int_0^\cdot q\theta_u\, du \tag{3.9}$$

is (locally) a $\bar{\mathbb{P}}$-Brownian motion and the dynamics of the process Y can be conveniently written as

$$dY_t = \alpha_q(\theta_t, \Sigma_t)\, dt + \Sigma_t\, d\bar{B}_t.$$

The expression inside the infimum in (3.8) involves a discounted running cost. Hence, it fits in the classical framework of optimal stochastic control, and a formal HJB-equation can be written down. We note that even though the process $\mathcal{E}(-\theta \cdot B)$ appears in the original expression for w, the simplification in (3.8) allows us to drop it from the list of state variables and, thus, reduce the dimensionality of the problem. Indeed, the formal HJB has the following form:

$$\inf_{\Sigma, \theta}\left(\tfrac{1}{2}\Sigma^2 w''(y) + \alpha_q(\Sigma, \theta)w'(y) - \beta(\theta)w(y) + \gamma(\theta)\right) = 0, \; y \in (\underline{y}, \overline{y}) \tag{3.10}$$

where the functions β and γ are defined in (2.11).

In order to fully characterize the optimization problem, we need to impose the boundary conditions at $\underline{y}$ and $\overline{y}$ to enforce the requirement that Y stay within the interval $[\underline{y}, \overline{y}]$. These amount to turning off the diffusion completely and leaving only the drift in the appropriate (inward) direction

---

[1]One should rather call $\bar{\mathbb{P}}$ a cylindrical measure, but, given the heuristic nature of the present section, we do not pursue this distinction.

when Y reaches the boundary. Thanks to the form of the function α_q and the equation (3.10), as well as the expectation that w' be bounded on $[\underline{y}, \overline{y}]$, we are led to the boundary condition

$$w''(\underline{y}) = w''(\overline{y}) = +\infty. \tag{3.11}$$

It will be shown in the following section that, in addition to the annihilation of the diffusion coefficient, (3.11) will ensure that the drift coefficient α_q will indeed have the proper sign of at the boundary.

3.2. **Shadow price as a strategy in a game.** By interpreting the problem of shadow prices as a game, one can arrive to the two-point boundary problem (3.10), (3.11) without the use of duality.

Let $\xi(y) = \eta_B + \eta_S S_0 e^y$ denote the initial wealth of the investor who is not subject to transaction costs, for a fixed η_B, η_S and $y \in [\overline{y}, \underline{y}]$. Finding a shadow price now amounts to the following:

(1) given (θ, Σ), solving the optimal investment problem for an investor with initial wealth $\xi(y)$ i.e., maximizing the expected discounted utility

$$\mathbb{E}[\int_0^\infty e^{-\delta t} U(c_t)].$$

over (φ, c) (numbers of shares of $\tilde{S}$ and the rate of consumption).
(2) minimizing the obtained value over (θ, Σ), and
(3) finally, further minimizing over $y \in [\overline{y}, \underline{y}]$.

Up to the last minimization over y, the above defines a stochastic game with the value

$$v(\xi, y) := \inf_{\theta, \Sigma} \sup_{\pi, c} \mathbb{E}[\int_0^\infty e^{-\delta t} U(c_t)].$$

The corresponding Isaacs equation with a two dimensional state (Ξ, Y) and the initial condition $(\Xi_0, Y_0) = (\xi, y)$ scales as

$$v(\xi, y) = \frac{\xi^p}{p} |w(y)|^{1-p} \text{ (compare to (3.6))}.$$

Consequently, it can be reduced to a one-dimensional equation for $w(y)$, with $y \in [\overline{y}, \underline{y}]$, which turns out to be exactly the boundary-value problem (3.10), (3.11). Once the game is solved, we can find a shadow price by simply minimizing the value $v(\xi(y), y)$ over $y \in [\overline{y}, \underline{y}]$.

We believe that a similar approach - namely of rewriting the problem of optimal investment and consumption with transaction costs as a game, through the use of consistent prices - works well in more general situations, e.g., when multiple assets are present.

3.3. **An order reduction.** Finally, based on the fact that the equation (2.10) is autonomous, we introduce an order-reducing change of variable. With w' expected to be increasing and continuous on $[\underline{y}, \overline{y}]$, we define the function $g : [\underline{x}, \overline{x}] \to \mathbb{R}$, with $\overline{x} = -w'(\underline{y})$ and $\underline{x} = -w'(\overline{y})$ by $w(y) = g(-w'(y))$. This transforms the equation (2.10) into

$$\inf_{\Sigma, \theta} \left(\tfrac{1}{2} \Sigma^2 \tfrac{x}{g'(x)} - \alpha_q(\Sigma, \theta) x - \beta(\theta) g(x) + \gamma(\theta) \right) = 0,$$

with (free) boundary conditions $g'(\underline{x}) = g'(\overline{x}) = 0$ and $\int_{\underline{x}}^{\overline{x}} \frac{g'(x)}{x} dx = \overline{y} - \underline{y}$. The free boundaries $\underline{x}$ and $\overline{x}$ are expected to be positive.

4. Proof of the main theorem: verification

We start the proof of our main Theorem 2.8 with a verification argument which establishes the implication (1) $\implies$ (2). After that, in Lemma 4.4 and Proposition 4.5, we show (3).

Let us assume, therefore, that a triplet $(\underline{x}, \overline{x}, g)$, as in part (1) of Theorem 2.8, is given (and fixed for the remainder of the section), and that the function h is defined as in (2.13). Let $\hat{\theta} : [\underline{x}, \overline{x}] \to \mathbb{R}$ and $\hat{\Sigma} : [\underline{x}, \overline{x}] \to \mathbb{R}$ be the formal optimizers of (2.10), i.e.,

$$\begin{cases} \hat{\theta}(x) = \frac{\sigma x}{h(x)}, \ \hat{\Sigma}(x) = -\frac{\sigma(1-x)g'(x)}{h(x)}, & p = 0 \\ \hat{\theta}(x) = -\frac{\sigma(1-p)x(qg'(x)-1)}{h(x)}, \ \hat{\Sigma}(x) = -\frac{\sigma(qg(x)-x)g'(x)}{h(x)}, & p \neq 0. \end{cases} \tag{4.1}$$

Similarly, let $\hat{\alpha}_q, \hat{\alpha}_0, \hat{\beta}, \hat{\gamma} : [\underline{x}, \overline{x}] \to \mathbb{R}$ be the compositions of the functions α_q, α_0, β and γ of (2.11) with $\hat{\theta}$ and $\hat{\Sigma}$. Using the explicit formulas above, one readily checks that function $\hat{\Gamma}(x) = -\frac{x}{g'(x)}\hat{\Sigma}(x)$ admits a Lipschitz extension to $[\underline{x}, \overline{x}]$.

While the equation (2.10) can be written in a more explicit way - which will be used extensively later - for now we choose to keep its current variational form. We do note, however, the following useful property of the function g:

Proposition 4.1. *For all $x \in (\underline{x}, \overline{x})$ with $g'(x) \neq 0$, we have*

$$\tfrac{1}{2}\hat{\Sigma}^2(x)\tfrac{d}{dx}\left(\tfrac{x}{g'(x)}\right) - \hat{\alpha}_q(x) - g'(x)\hat{\beta}(x) = 0. \tag{4.2}$$

Proof. The equation (4.2) follows either by direct computation (using the explicit formulas (4.1) for $\hat{\Sigma}$ and $\hat{\theta}$ above) or the appropriate version of the Envelope Theorem (see, e.g., Theorem 3.3, p. 475 in [GK02]), which states, loosely speaking, that we "pass the derivative inside the infimum" in the equation (2.10). $\square$

4.1. Construction of the state processes.
The family of processes $\{X_t^x\}_{t\in[0,\infty)}$, $x \in [\underline{x}, \overline{x}]$, defined in this section, will play the role of state processes in the construction of the shadow-price process below. Thanks to the Lipschitz property of the function h, for each $x \in [\underline{x}, \overline{x}]$ there exists a unique solution $\left(\{X_t^x\}_{t\in[0,\infty)}, \{\Phi_t^x\}_{t\in[0,\infty)}\right)$ of the following reflected (Skorokhod-type) SDE

$$\begin{cases} dX_t^x = \left(X_t^x\hat{\beta}(X_t^x) - q\hat{\theta}(X_t^x)\hat{\Gamma}(X_t^x)\right)dt + \hat{\Gamma}(X_t^x)\,dB_t + d\Phi_t^x, \\ X_0^x = x \in [\underline{x}, \overline{x}]. \end{cases} \tag{4.3}$$

Here, Φ is the "instantaneous inward reflection" term for the boundary $\{\underline{x}, \overline{x}\}$, i.e., a continuous process of finite variation whose pathwise Hahn-Jordan decomposition $(\Phi^{x\uparrow}, \Phi^{x\downarrow})$ satisfies

$$\Phi_t^{x\uparrow} = \int_0^t \mathbf{1}_{\{X_u^x = \underline{x}\}}\,d\Phi_u^{x\uparrow}, \text{ and } \Phi_t^{x\downarrow} = \int_0^t \mathbf{1}_{\{X_u^x = \overline{x}\}}\,d\Phi_u^{x\downarrow}, \ t \geq 0.$$

The reader is referred to [Sko61] for a more detailed discussion of various possible boundary behaviors of diffusions in a bounded interval, as well as the original existence and uniqueness result [Sko61, pp. 269-274]) for (4.3).

For $x \in [\underline{x}, \overline{x}]$, we define the function $f : [\underline{x}, \overline{x}] \to \mathbb{R}$ by $f(x) = \underline{y} + \int_x^{\overline{x}} \frac{g'(\xi)}{\xi}\,d\xi$ and the process $\{Y_t^x\}_{t\in[0,\infty)}$ by $Y_t^x = f(X_t^x)$. In relation to the heuristic discussion of Section 3, we note that f plays the (formal) role of the inverse of the derivative w'. Moreover, the process Y^x has the following properties:

Proposition 4.2. *For $x \in [\underline{x}, \overline{x}]$, we have*

(1) $Y_t^x \in [\underline{y}, \overline{y}]$, for all $t \geq 0$, a.s., and

(2) $Y_0^x = f(x)$ *and* $dY_t^x = \hat{\alpha}_0(X_t^x)\, dt + \hat{\Sigma}(X_t^x)\, dB_t$.

Proof. Property (1) follows from the definition of the function f and the assumption (c) of part (1) of Theorem 2.8. For (2), Itô's formula reveals the following dynamics of Y^x:

$$dY_t^x = \left(-g'(X_t^x)\hat{\beta}(X_t^x) - q\hat{\Sigma}(X_t^x)\hat{\theta}(X_t^x) + \tfrac{1}{2}\hat{\Sigma}^2(X_t^x)\tfrac{d}{dx}\left(\tfrac{x}{g'(x)}\right)\Big|_{x=X_t^x}\right) dt + \hat{\Sigma}(X_t^x)\, dB_t - \tfrac{g'(X_t^x)}{X_t^x}\, d\Phi_t^x.$$

The identity (4.2) allows us to simplify the above expression to

$$dY_t^x = \hat{\alpha}_0(X_t^x)\, dt + \hat{\Sigma}(X_t^x)\, dB_t - \tfrac{g'(X_t^x)}{X_t^x}\, d\Phi_t^x.$$

Finally, since g' vanishes on the boundary, the singular term disappears and we obtain the second statement. $\square$

4.2. A stochastic representation for the function g.

For notational convenience, we define

$$W_t^x = \begin{cases} -\delta \log\left(\mathcal{E}\left(-\hat{\theta}(X^x)\cdot B\right)_t\right), & p = 0, \\ \operatorname{sgn}(p)\mathcal{E}(-\hat{\theta}(X^x)\cdot B)_t^{-q}, & p \neq 0. \end{cases}$$

Proposition 4.3. *For $x \in [\underline{x}, \overline{x}]$, we have*

$$g(x) = \mathbb{E}\Big[\int_0^\infty e^{-\hat{\delta}t} W_t^x\, dt\Big]. \tag{4.4}$$

Proof. Using the equation (2.10), relation (4.2) and Itô's formula, we can derive the following dynamics for the process $g(X_t^x)$:

$$dg(X_t^x) = \left(\hat{\beta}(X_t^x)g(X_t^x) - \hat{\gamma}(X_t^x)\right) dt + g'(X_t^x)\hat{\Gamma}(X_t^x)\, d\bar{B}_t$$

where $\bar{B}$ is given by (3.9) with $\theta_t = \hat{\theta}(X_t^x)$. On the other hand, if we set $\rho_t^x = e^{-\int_0^t \hat{\beta}(X_u^x)\, du}$ and $H_t^x = \int_0^t \rho_u^x \hat{\gamma}(X_u^x)\, du + \rho_t^x g(X_t^x)$, we obtain that

$$dH_t^x = \rho_t^x g'(X_t^x)\hat{\Gamma}(X_t^x)\, d\bar{B}_t.$$

Girsanov's theorem (applicable thanks to the boundedness of $\hat{\theta}$) implies that $\bar{B}$ is a Brownian motion on $[0, t]$, under the measure $\bar{\mathbb{P}}_t$, defined by $d\bar{\mathbb{P}}_t = \mathcal{E}(q\hat{\theta}\cdot B)_t\, d\mathbb{P}$. Therefore,

$$\mathbb{E}^{\bar{\mathbb{P}}_t}[\rho_t^x g(X_t^x)] + \mathbb{E}^{\bar{\mathbb{P}}_t}\left[\int_0^t \rho_u^x \hat{\gamma}(X_u^x)\, du\right] = \mathbb{E}^{\bar{\mathbb{P}}_t}[H_t] = \mathbb{E}^{\bar{\mathbb{P}}_t}[H_0] = g(x),$$

where the boundedness of the integrands was used to do away with the stochastic integrals with respect to $\bar{B}$. The exponential identity (3.7) now implies that

$$g(x) = \mathbb{E}^{\bar{\mathbb{P}}_t}[\rho_t^x g(X_t^x)] + \mathbb{E}\Big[\int_0^t e^{-\hat{\delta}u} W_u^x\, du\Big] + \begin{cases} 0, & p \neq 0 \\ \tfrac{1}{2}e^{-\delta t}\mathbb{E}[\int_0^t \hat{\theta}(X_s^x)^2\, ds], & p = 0 \end{cases} \tag{4.5}$$

For $p = 0$, we can use the fact that $\hat{\theta}$ and g are bounded to conclude that

$$\mathbb{E}^{\bar{\mathbb{P}}_t}[\rho_t^x g(X_t^x)] = \mathbb{E}[e^{-\delta t} g(X_t^x)] \to 0 \quad \text{and} \quad e^{-\delta t}\mathbb{E}\Big[\int_0^t \hat{\theta}(X_s^x)^2\, ds\Big] \to 0.$$

These two limits can now easily be combined with (4.5) to yield (4.4).

To deal with the case $p > 0$, we note that non-negativity of g and W_t^x in (4.5) implies that

$$\int_0^\infty e^{-\hat{\delta}t}\mathbb{E}[W_t^x]\, dt < \infty. \tag{4.6}$$

Moreover, with $|g|_\infty = \sup_{x \in [\underline{x}, \overline{x}]} |g(x)|$, we have

$$\mathbb{E}^{\bar{\mathbb{P}}_t}[\rho_t^x g(X_t^x)] \leq |g|_\infty \mathbb{E}^{\bar{\mathbb{P}}_t}[\rho_t^x] = |g|_\infty \, e^{-\hat{\delta} t} \mathbb{E}[W_t^x].$$

Therefore, it is enough to observe that $e^{-\hat{\delta} t_n} \mathbb{E}[W_{t_n}^x] \to 0$, along sequence $\{t_n\}_{n \in \mathbb{N}}$ with $t_n \to \infty$ which exists thanks to (4.6).

For $p < 0$, the fact that $\rho_t^x \leq e^{-\hat{\delta} t}$ implies that $\mathbb{E}^{\bar{\mathbb{P}}_t}[\rho_t^x |g(X_t^x)|] \leq |g|_\infty \, e^{-\hat{\delta} t} \to 0$, which, in turn, together with (4.5), implies (4.4). $\qquad\square$

4.3. **The Shadow Market.** For $x \in [\underline{x}, \overline{x}]$, we define the process $\hat{S}_t^x = S_t e^{Y_t^x}$ and observe that, by Itô's formula, it admits the following dynamics:

$$d\hat{S}_t^x = \hat{S}_t^x \Big(\sigma + \hat{\Sigma}(X_t^x)\Big)\Big(\hat{\theta}(X_t^x)\, dt + dB_t\Big), \quad \hat{S}_0^x = S_0 e^{f(x)}. \tag{4.7}$$

The goal of this subsection is to show that $\hat{S}_t^x$ is a shadow price for the appropriate choice of the initial value $x \in [\underline{x}, \overline{x}]$. Note that $\sigma + \hat{\Sigma}(X_t^x)$ is bounded away from 0, by the properties of g.

Recall from subsection 2.5 that $\hat{u}(\hat{S}^x)$ is the value of the optimal consumption problem $\mathcal{U}(c) \to$ max in the (frictionless) financial market driven by $\hat{S}^x$ for an agent with the initial holding η_B in the bond, and η_S in the stock (the $\hat{S}^x$-problem). The following lemma, which describes the optimal investment/consumption policy that achieves the maximum, will play a key role in the proof of the shadow property of the process $\hat{S}^x$. To simplify the notation, we introduce the following shortcuts:

$$\xi(x) = \eta_B + S_0 e^{f(x)} \eta_S, \;\; \Pi(x) = \begin{cases} x, & p = 0, \\ \frac{x}{qg(x)}, & p \neq 0, \end{cases} \;\; \text{and} \;\; K(x) = \begin{cases} \delta & p = 0, \\ \frac{1}{|g(x)|}, & p \neq 0. \end{cases}$$

Lemma 4.4. *For $x \in [\underline{x}, \overline{x}]$ and the initial positions (η_B, η_S) with $\eta_B + S_0 e^{f(x)} \eta_S \geq 0$, we have*

$$\hat{u}(\eta_B, \eta_S; x) = \begin{cases} \frac{1}{\delta}\Big(-1 + \log(\delta \xi(x)) + g(x)\Big), & p = 0, \\ \frac{1}{p} \xi(x)^p \, |g(x)|^{1-p}, & p \neq 0. \end{cases} \tag{4.8}$$

Moreover, with the processes $\{\hat{\pi}_t^x\}_{t \in [0,\infty)}$, $\{\hat{\kappa}_t^x\}_{t \in [0,\infty)}$ and $\{\hat{V}_t^x\}_{t \in [0,\infty)}$ defined by

$$\hat{\pi}_t^x = \Pi(X_t^x), \quad \hat{\kappa}_t^x = K(X_t^x), \quad \hat{V}_t^x = \xi(x)\mathcal{E}\Big(\int_0^\cdot \frac{\hat{\pi}_u^x}{\hat{S}_u^x} d\hat{S}_u^x - \int_0^\cdot \hat{\kappa}_u^x \, du\Big)_t, \tag{4.9}$$

the optimal strategy $(\hat{\varphi}^{0,x}, \hat{\varphi}^x, \hat{c}^x)$ for the $\hat{S}^x$-problem is given for $t \geq 0$ by

$$\hat{c}_t^x = \hat{V}_t^x \hat{\kappa}_t^x, \qquad \hat{\varphi}_t^{0,x} = \hat{V}_t^x (1 - \hat{\pi}_t^x) \quad \text{and} \quad \hat{\varphi}_t^x = \frac{\hat{V}_t^x \hat{\pi}_t^x}{\hat{S}_t^x}. \tag{4.10}$$

Proof. The standard complete-market duality theory (see, e.g., Theorem 9.11, p. 141 in [KS98]) implies that

$$\hat{u}(\eta_B, \eta_S; x) = \inf_{z > 0} \Big((\eta_B + S_0 e^{f(x)} \eta_S) z + \mathcal{V}\Big(z\mathcal{E}\big(-\hat{\theta}(X^x) \cdot B\big)\Big)\Big)$$

where dual functional $\mathcal{V}$ is as in (3.4). Furthermore, following the computations that lead to (3.6) in subsection 3.1, and using the representation of Proposition 4.3, we get (4.8)

Once the form of the value function $\hat{u}$ has been determined, it is a routine computation derive the expressions for the optimal investment/consumption strategy. Indeed, let the processes $\{\hat{\pi}_t^x\}_{t \in [0,\infty)}$ be given by

$$\hat{\pi}_t^x = \frac{\hat{\theta}(X_t^x)}{(1-p)\big(\sigma + \hat{\Sigma}(X_t^x)\big)} + \begin{cases} 0, & p = 0, \\ -\frac{X_t^x}{g(X_t^x)} \frac{\hat{\Sigma}(X_t^x)}{\sigma + \hat{\Sigma}(X_t^x)}, & p \neq 0, \end{cases} \tag{4.11}$$

and $\{\hat{\kappa}_t^x\}_{t \in [0,\infty)}$ and $\{\hat{V}_t^x\}_{t \in [0,\infty)}$ as in the statement. Then, one readily checks that the triplet $(\hat{\varphi}^{0,x}, \hat{\varphi}^x, \hat{c}^x)$ given by (4.10) is an optimal investment/consumption strategy.

Finally, the equality between the form (4.11) and the simpler one given in (4.9) in the statement follows by direct computation where one can use the explicit formulas for the functions $\hat{\Sigma}$ and $\hat{\theta}$ from (4.1).   □

Proposition 4.5. *Let (η_B, η_S) be an admissible initial wealth, i.e., such that $\mathrm{Liq}(\eta_B, \eta_S, (1 - \underline{\lambda})S_0, (1+\overline{\lambda})S_0) \geq 0$. For the function $r : [\underline{x}, \overline{x}] \to \mathbb{R}$, given by $r(x) = \eta_S S_0 e^{f(x)}(1 - \Pi(x)) - \eta_B \Pi(x)$, let the constant $\hat{x} \in [\underline{x}, \overline{x}]$ be defined by*

$$
\hat{x} = \begin{cases}
\overline{x}, & r(x) > 0 \text{ for all } x \in [\underline{x}, \overline{x}] \\
\underline{x}, & r(x) < 0 \text{ for all } x \in [\underline{x}, \overline{x}] \\
a \text{ solution to } r(x) = 0, & otherwise.
\end{cases}
$$

Then $\hat{S} = \hat{S}^{\hat{x}}$ is a shadow price.

Remark 4.6. The three possible cases in Proposition 4.5 relate to whether the initial condition is outside the no-transaction region (above or below) or inside it. It is easy to check that $\hat{x}$ minimizes the value $\frac{\xi(x)^p}{p}|g(x)|^{1-p}$ for $\xi(x) = \eta_B + \eta_S S_0 e^{f(x)}$, as mentioned in Subsection 3.2.

Proof. The idea of the proof is to show that the triplet $(\hat{\varphi}^0, \hat{\varphi}, \hat{c})$ of Lemma 4.4 satisfies the conditions of Proposition 2.4. Since $\hat{c}$ is the optimal consumption process, it will be enough to show that conditions (2) and (3) of Proposition 2.2 hold. The expression (4.10) implies that the processes $\hat{\varphi}^0$ and $\hat{\varphi}$ are continuous, except for a possible jump at $t = 0$.

Let us, first, deal with the jump at $t = 0$. The conditions (2) and (3) of Proposition 2.2 at $t = 0$ translate into the following equality:

$$
\hat{\varphi}_{0+}^0 - \eta_B + (1 + \overline{\lambda})S_0(\varphi_{0+} - \eta_S)^+ - (1 - \underline{\lambda})S_0(\varphi_{0+} - \eta_S)^- = 0,
$$

which, after (4.10) is used, becomes

$$
e^{f(x)}r(x) = G(r(x)), \text{ where } G(x) = (1 - \underline{\lambda})x^+ - (1 + \overline{\lambda})x^-. \tag{4.12}
$$

If $r(x) = 0$ admits a solution $\hat{x} \in [\underline{x}, \overline{x}]$, then $x = \hat{x}$ clearly satisfies the equation (4.12). On the other hand, if $r(x) \neq 0$, for all $x \in [\underline{x}, \overline{x}]$, then by continuity, either $r(x) > 0$, for all $x \in [\underline{x}, \overline{x}]$ or $r(x) < 0$, for all $x \in [\underline{x}, \overline{x}]$. Focusing on the first possibility (with the second one being similar) we note that in this case $G(r(x)) = (1 - \underline{\lambda})r(x)$, and so, if we pick $\hat{x} = \overline{x}$, we get $e^{f(\hat{x})}r(\hat{x}) = (1 - \underline{\lambda})r(\hat{x}) = G(r(\hat{x}))$.

Next, we deal with the trajectories of the processes $\hat{\varphi}^0$ and $\hat{\varphi}$ for $t > 0$. It is a matter of a tedious but entirely straightforward computation (which can be somewhat simplified by passing to the logarithmic scale and using the identities (2.10) and (4.2)) to obtain the following dynamics:

$$
d\hat{\varphi}_t = \frac{\hat{\varphi}_t}{X_t^x} d\Phi_t^x. \tag{4.13}
$$

Thanks to the fact that Φ^x is a finite-variation process which decreases only when $X_t^x = \overline{x}$ (i.e., $\hat{S}_t^x = \underline{S}$) and increases only when $X_t^x = \underline{x}$ (i.e., $\hat{S}_t^x = \overline{S}$), the conditions (2) and (3) of Proposition 2.2 hold.   □

5. Main ideas behind the proof of existence for the free-boundary problem

Having presented a verification argument in the previous section, we turn to the analysis of the (non-standard) free-boundary problem (2.10), (2.12). We start by remarking that that the equation (2.10) simplifies to the form

$$g'(x) = L(x, g(x)), \text{ where } L(x, z) = \frac{P(x, z)}{Q(x, z)},$$

and where the second-order polynomials $P(x, z)$ and $Q(x, z)$ are given by

$$P(x, z) = \begin{cases} -2q\delta z^2 + 2p(\mu x + \operatorname{sgn}(p))z - (1-p)^2 \sigma^2 x^2, & p \neq 0 \\ -2\delta z - \sigma^2 x^2 + 2\mu x, & p = 0 \end{cases}$$

$$Q(x, z) = \begin{cases} -P(x, z) + (p\sigma^2 - 2\hat{\delta})xz + 2(\mu - (1-p)\sigma^2)x^2 + 2\operatorname{sgn}(p)x, & p \neq 0 \\ (1-x)(2\delta z + (\sigma^2 - 2\mu)x), & p = 0 \end{cases}$$

The existence proof is based on a geometrically-flavored analysis of the equation (2.10), where the curves $\mathbb{T}$ and $\mathbb{B}$, given by

$$\begin{aligned} \mathbb{T} &= \left\{ (x, z) \in (0, \infty) \times \mathbb{R} : P(x, z) = 0 \right\} \\ \mathbb{B} &= \left\{ (x, z) \in (0, \infty) \times \mathbb{R} : Q(x, z) = 0 \right\}, \end{aligned} \tag{5.1}$$

play a prominent role. Many cases need to be considered, but we always proceed according to the following **program**:

(1) First, we note that the boundary conditions $g'(\underline{x}) = g'(\overline{x}) = 0$ amount to

$$(\underline{x}, g(\underline{x})), (\overline{x}, g(\overline{x})) \in \mathbb{T}.$$

(2) Then, for a fixed $(\alpha, z(\alpha)) \in \mathbb{T}$ we solve the ODE $g'(x) = L(x, g(x))$ with initial condition $g(\alpha) = z(\alpha)$ and let it evolve to the right (if possible) until meeting again the curve $\mathbb{T}$ at the x-intercept β_α. We therefore obtain a solution $g_\alpha : [\alpha, \beta_\alpha] \to \mathbb{R}$ satisfying

$$g_\alpha'(\alpha) = g_\alpha'(\beta_\alpha) = 0.$$

If $P = Q = 0$ on some point along the way, only continuity is required there.

(3) Finally, we vary the parameter α to meet the integral condition $\int_{\underline{x}}^{\overline{x}} \frac{g'(x)}{x} \, dx = \log(\frac{1+\overline{\lambda}}{1-\underline{\lambda}})$.

In order to give some intuition for the technicalities that follow, let us consider, for a moment, the "degenerate" frictionless case $\underline{\lambda} = \overline{\lambda} = 0$. For fixed μ, σ, p, and δ, the absence of transaction costs suggests a trivial solution with $\underline{x} = \overline{x}$. In addition, the point $(\overline{x}, g(\overline{x})) \in \mathbb{T}$ is expected to have the highest possible z-coordinate. Indeed, larger values of g translate, as we saw in Lemma 4.4, to larger expected utilities. If such a point exists we call it the **North pole**, denote it by N and its x-coordinate by $x_N > 0$. In that case, furthermore, the curve $\mathbb{T}$ decomposes into two parts $\mathbb{W}$ (West of North) and $\mathbb{E}$ (East of North) so that

$$\mathbb{T} = \mathbb{W} \cup \{N\} \cup \mathbb{E}.$$

Remark 5.1. It turns out that:
(1) The curve $\mathbb{T}$ has a North pole, if and only if $u < \infty$ when $\underline{\lambda} = \overline{\lambda} = 0$.
(2) When the North pole does exist:
   (a) if $\underline{\lambda} = \overline{\lambda} = 0$ then $\underline{x} = \overline{x} = x_N$ and $(\underline{x}, g(\underline{x})) = N$, and
   (b) if $\underline{\lambda} + \overline{\lambda} > 0$, we expect $(\underline{x}, g(\underline{x})) \in \mathbb{W}$ and $(\overline{x}, g(\overline{x})) \in \mathbb{E}$.

Before we go ahead, we note that the quantities $\frac{2\delta}{p}$ and $(1-p)\sigma^2$ together with their geometric and arithmetic means play a special role. In fact, they deserve their own notation:

$$G = G(\sigma, p, \delta) = \sqrt{\frac{2\delta(1-p)\sigma^2}{p}}, \quad A = A(\sigma, p, \delta) = \frac{\delta}{p} + \frac{(1-p)\sigma^2}{2}.$$

Another quantity that will play a role in the analysis is the *Merton proportion*

$$\pi = \pi(\mu, \sigma, p) = \frac{\mu}{(1-p)\sigma^2},$$

for an investor in a frictionless market, with the power utility. The last thing we need to do before we delve deeper into the analysis of various cases, is to introduce a suitable notation for the **singular** points, i.e., the points $(x, z) \in (0, \infty) \times \mathbb{R}$ with $P(x, z) = Q(x, z) = 0$. The explicit expressions for P and Q above yield immediately that, in general, there are three($p \neq 0$) or two($p = 0$) solutions to $P = Q = 0$ in $\mathbb{R}^2$, two($p \neq 0$) or one($p = 0$) of which lie on the z axis (and, therefore, do not count as singular points). The other one, denoted by P is the unique singular point and will be quite important in our analysis. It has coordinates

$$x_P = \begin{cases} \frac{\text{sgn}(p)}{A-\mu}, & p \neq 0, \\ 1, & p = 0, \end{cases} \quad \text{and} \quad z_P = \begin{cases} \frac{1}{q}x_P, & p \neq 0, \\ \frac{2\mu-\sigma^2}{2\delta}, & p = 0, \end{cases}$$

which clearly degenerate for $A \leq \mu$, $p \neq 0$; in those cases, we set $(x_P, z_P) = (\infty, \infty)$.

We are now ready to start differentiating between several (technically different) cases which are chosen, roughly, according to the following criteria: (1) whether the risk aversion is *low* $(0 < p < 1)$ or *high* $(p \leq 0)$, (2) whether the "North pole" exists, and (3) the sign of $\pi - 1$.

5.1. **Low risk aversion** $0 < p < 1$. In this case investor is less risk averse than the log-investor, and it is the only case when well-posedness may fail. We further separate it into several sub-cases:

- *Sub-case a):* $\mu < G$. For these particular values of parameters, the problem turns out to be well posed. The reason is simple: the value function of the *frictionless* version is finitely-valued here. The curve $\mathbb{T}$ is (a portion of) an ellipse, and, as such, it obviously has a "North pole", in agreement with Remark 5.1.

Let E denote the most right-ward point (East) and by x_E its x-coordinate, so that $0 < x_N < x_E$. Taking into account Remark 5.1 and the fact that $\mathbb{T}$ is an ellipse, we expect to find a solution $(\underline{x}, \overline{x}, g)$ of the free boundary which satisfies $\underline{x} < \overline{x} \leq x_E$. As described in the outline of our program above, we "evolve" the solution from the initial point to the right, as long as we can. More precisely, we consider a maximal (with respect to the domain) C^2-solution of the initial-value problem $g'_\alpha(x) = L(x, g_\alpha(x))$, $g'(\alpha) = 0$ with the property that $P(x, g_\alpha(x)) \geq 0$, i.e., such that the curve g_α stays on the inside of $\mathbb{T}$. It turns out that the domain of this solution is of the form $[\alpha, \beta_\alpha]$, for some $\beta_\alpha \in [\alpha, x_E]$, and that the following statements hold:

(1) The map $\alpha \mapsto \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx$ is continuous and strictly decreasing on $(0, x_N)$, and
(2) $\lim_{\alpha \searrow 0} \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx = \infty$ while $\lim_{\alpha \nearrow x_N} \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx = 0$.

It follows immediately that a unique α, such that g_α solves the free-boundary problem (2.10),(2.12) exists. The major difficulty in the analysis is the fact that, for a given α, the maximal solution g_α may encounter the singularity P on its trajectory (see Figure 3. below).

An important tool here turns out the be the so-called **containment curve**, i.e., a function $\tau : (0, \infty) \to \mathbb{R}$ such that

- g_α cannot hit τ before it hits $\mathbb{T}$, and
- g_α must hit τ before it hits $\mathbb{B}$.

It serves a two-fold purpose here. First, it restricts the possible values the function g_α can take and makes sure that it either does not intersect the (singular) curve $\mathbb{B}$ at all, or that it encounters it only at the point P. The shaded area Ω_0 in Figure 2. below depicts the region of the plane the graph Γ_α is restricted to lie in, under various conditions on the problem parameters. Second, when the singular point P indeed happens to lie on the graph Γ_α, a well-constructed containment curve τ provides crucial information about the behavior of g_α in a neighborhood of P. Whether or not P falls on the graph

$$\Gamma_\alpha = \{(x, g_\alpha(x)) \, : \, x \in [\alpha, \beta_\alpha]\}.$$

of g_α depends on the values of the parameters. In particular, it depends on the relative position of the points E, P and N. The lead actor turns out to be the Merton proportion $\pi = \pi(\mu, \sigma, p)$, and the following three cases need to be distinguished (see Figure 2., below):

(1) $\pi < 1$: $P \in \mathbb{E}$ and $P \notin \Gamma_\alpha$, with the relative positions of P, E and N, further determined by the sign of $\frac{2\delta}{p} - (1-p)\sigma^2$

(2) $\pi = 1$: Here, $P = N$ and $\beta_\alpha = x_P$ for any α.

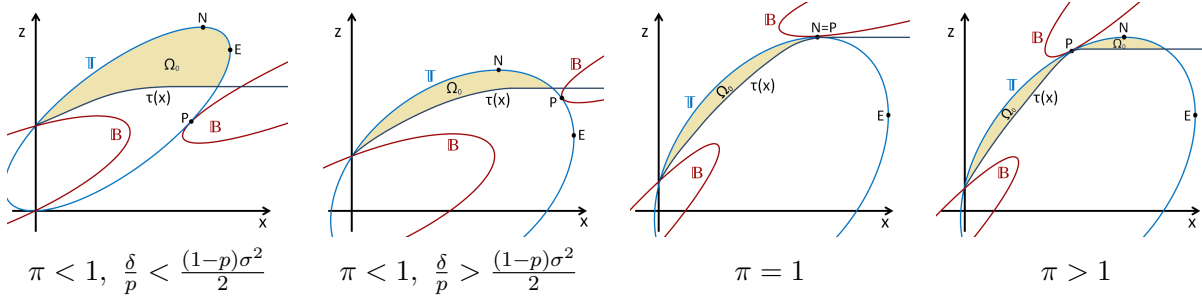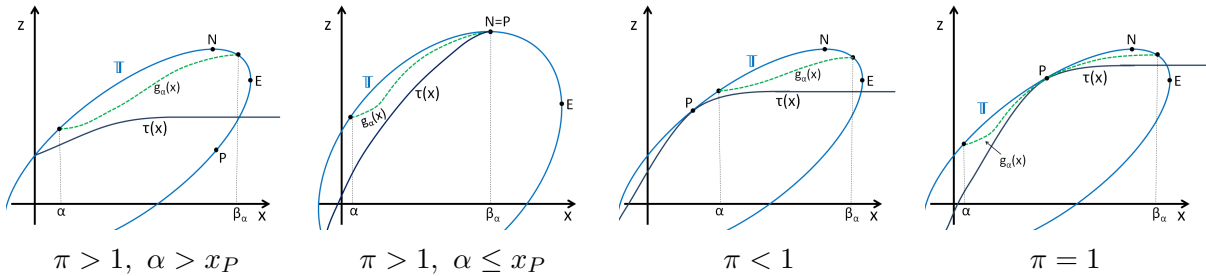(3) $\pi > 1$: In this case, $P \in \mathbb{W}$. Furthermore, $P \in \Gamma_\alpha$ if and only if $\alpha \leq x_P$.

$$\pi < 1, \ \frac{\delta}{p} < \frac{(1-p)\sigma^2}{2} \qquad \pi < 1, \ \frac{\delta}{p} > \frac{(1-p)\sigma^2}{2} \qquad \pi = 1 \qquad \pi > 1$$

Figure 2. $0 < p < 1, \ \mu < G$.

Figure 3. below shows some of the possible shapes the graph Γ_α can take, under a representative choice of parameter regimes.

$$\pi > 1, \ \alpha > x_P \qquad \pi > 1, \ \alpha \leq x_P \qquad \pi < 1 \qquad \pi = 1$$

Figure 3. $0 < p < 1, \ \mu < G$

A rigorous treatment of the first possibility ($\pi < 1$) is given in the Proposition 6.7 in Section 6. The other cases are treated in the Proposition 6.9.

- *Sub-case b): $\mu \geq A$.* The rate of return in this sub-case is so large, that the value function of the problem with transaction costs is infinite, independently of the size of the transaction costs $\underline{\lambda}$ and $\overline{\lambda}$. A constructive argument is presented in Proposition 6.1. From the analytic point of view, this phenomenon is related to the non-existence of the solution to the free-boundary problem; an illustration of the reason why is given by the picture to the right (Figure 4). In a nutshell, we can find an asymptotically linearly increasing curve $T_u(x, K)$ (the notation is chosen to fit that of Section 6) such that g_α stays above it for all x. Consequently, it is prevented from reaching the other branch of the curve $\mathbb{T}$ and satisfying the free-boundary condition.
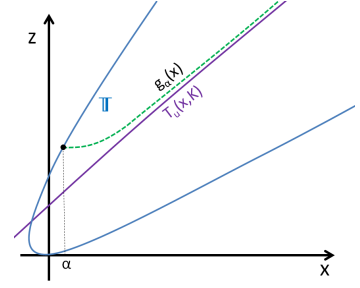
Figure 4.

- *Sub-case c): $G \leq \mu < A$.* This is the most interesting sub-case from the point of view of well-posedness; whether the value function is finite or not is determined by the size of the transaction costs. The curve $\mathbb{T}$ is a hyperbola, and has no North pole. The overall approach is the same as in sub-case a): we construct a maximal solution g_α on an interval of the form $[\alpha, \beta_\alpha]$, and show that the following two statements hold:

(1) The map $\alpha \mapsto \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx$ is continuous and strictly decreasing on $(0, \infty)$, and

(2) $\lim_{\alpha \searrow 0} \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx = \infty$ while $\lim_{\alpha \nearrow \infty} \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx = C$,

where an expression for $C = C(\mu, \sigma, \delta, p)$ is given in (6.7) below. The reader will note two major differences when the statements here are compared to the corresponding statements in the sub-case a). The first one is that $+\infty$ now plays the role of x_N. The second one is that the range of the integral $\int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx$ is not the set of positive numbers anymore. It is an interval of the form (C, ∞), which makes the free-boundary problem solvable only for $\log(\frac{1+\overline{\lambda}}{1-\underline{\lambda}}) > C$.

In addition to the fact that we still need to deal with the possible singularity along the graph Γ_α of g_α, difficulties of a different nature appear in this sub-case. First of all, due to the unboundedness of the regions separated by a hyperbola, it is not clear whether the maximal solution started at $x = \alpha$ will ever hit the curve $\mathbb{T}$ again. Indeed, this is certainly a possibility when $\mu \geq A$, as depicted in Figure 4. However, we prove by contradiction that this is not the case for $G \leq \mu < A$. The second new difficulty has to do with fact that C is finite - a fact which prevents the existence of a solution to (2.10), (2.12).

$$\pi < 1 \qquad\qquad \pi > 1, \ \alpha > x_P \qquad\qquad \pi > 1, \ \alpha \leq x_P$$
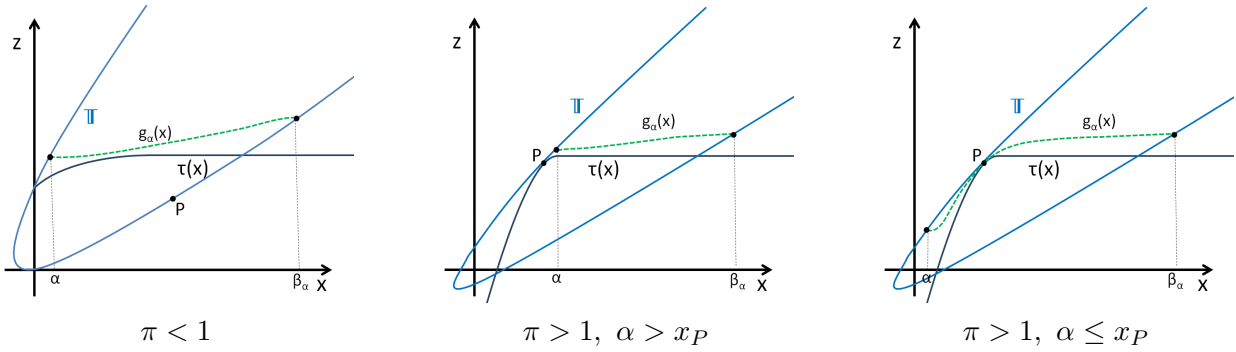
Figure 5. $0 < p < 1, \ G \leq \mu < A$

The rigorous treatment of this sub-case is in Propositions 6.13 and 6.14 in Section 6. Figure 5. below illustrates three representative regimes. We note that the equality $\pi = 1$ cannot hold for $\mu \in [A, G)$, as it would force $A = G$.

5.2. **High risk aversion $p \leq 0$.** In this case the problem is always well posed independently of the values of $\underline{\lambda}$ and $\overline{\lambda}$; indeed, the utility function is bounded from above. The curve $\mathbb{T}$ is a hyperbola for $p < 0$ and a parabola for $p = 0$, and it has a North-pole for any $p \leq 0$. $\mathbb{B}$ is a hyperbola for $p < 0$, and for $p = 0$, it is a union of two straight lines, one of which is $x = 1$.

Compared to the case $0 < p < 1$, no major new difficulties arise here, even though one still has to deal with the existence of singularities. For this reason we only present a figure (Figure 6. below) which illustrates different sub-cases that may arise. The formal treatment is analogous to that of Section 6.
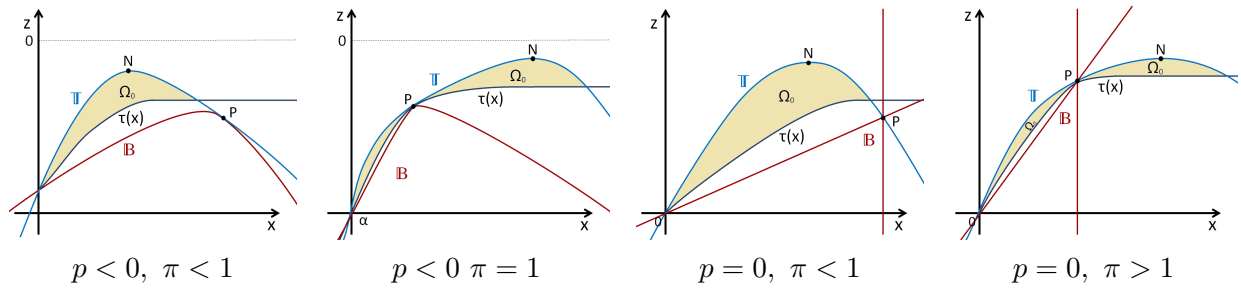

$$p < 0, \ \pi < 1 \qquad p < 0 \ \pi = 1 \qquad p = 0, \ \pi < 1 \qquad p = 0, \ \pi > 1$$

FIGURE 6. $p \leq 0$

6. AN EXISTENCE PROOF FOR THE FREE-BOUNDARY PROBLEM (2.10), (2.12)

After a heuristic description of the major steps in the existence proof and the associated difficulties, we now proceed to give more rigorous, formal proofs. More precisely, the goal of this section is to present a proof of the part (1) of Theorem 2.8.

As already mentioned in the previous section, the proofs in the case $p \leq 0$, are very similar (but less involved) than those in the case $p \in (0, 1)$ so we skip them and refer the reader to the first author's PhD dissertation [Cho12] for details. We also do not provide the proof of the part (c) of Theorem 2.8, as it can be obtained easily by an explicit computation.

Out first result states that problem is not well posed for large μ. As a consequence, we will be focusing on the case $0 < p < 1$, $\mu < A$ in the sequel.

Proposition 6.1 $(0 < p < 1, \mu \geq A.)$. *If $0 < p < 1$ and $\mu \geq A$, then $u = \infty$, for all $\underline{\lambda}, \overline{\lambda} \geq 0$.*

Proof. Without loss of generality, we consider the case $\eta_B = 0, \eta_S = 1$, and construct a portfolio (φ^0, φ, c) as follows:

$$\varphi_t^0 = 0, \ \varphi_t = (t+1)^{-\frac{1-p}{p}} \text{ and } c_t = \frac{1-p}{p}(1-\underline{\lambda}) S_t (t+1)^{-\frac{1}{p}}, \text{ for } t \geq 0.$$

One easily checks that it is admissible and that its expected utility is given by

$$\mathbb{E}\left[\int_0^\infty e^{-\delta t} \frac{c_t^p}{p} dt \right] = \frac{(1-p)^p (1-\underline{\lambda})^p}{p^{1+p}} \mathbb{E}\left[\int_0^\infty e^{-\delta t} \frac{S_t^p}{t+1} dt \right] = \frac{(1-p)^p (1-\underline{\lambda})^p}{p^{1+p}} \int_0^\infty e^{pt(\mu-A)} \frac{1}{t+1} dt = \infty. \quad \square$$

6.1. **Maximal inner solutions of** $g' = L(\cdot, g)$**.** As explained in the previous section, the main technique we employ in all of our existence proofs is the construction of a family of solutions to the equation $g' = L(\cdot, g)$, followed by the choice of the one that satisfies the appropriate integral condition. We, therefore, take some time here to define the appropriate notion of a solution to a singular ODE $g' = L(\cdot, g)$:

Definition 6.2. Let $\mathcal{D}$ be a convex interval in $(0, \infty)$. We say that a function $g : \mathcal{D} \to \mathbb{R}$ is a **continuous solution** of the equation $g' = L(\cdot, g)$ if

(1) g is continuous on $\mathcal{D}$,
(2) g is differentiable at x and $g'(x) = L(x, g(x))$, for all $x \in \operatorname{Int} \mathcal{D} \setminus \{x_P\}$

We note that any function with a single-point domain $\mathcal{D} = \{x\}$ is considered a continuous solution according to the above definition.

Remembering that $p \in (0, 1)$ and using the notation of the previous section we remark that the level curves $L = k$ are ellipses or hyperbolas, and, as such, they are not graphs in general.

We therefore introduce the **upper graph** $T_u(x, k)$ and the **lower graph** $T_d(x, k)$ of the level curve $L = k$ by

$$T_u(x, k) = \max\{z \in \mathbb{R} : P(x, z) = k\, Q(x, z)\}$$

and

$$T_d(x, k) = \min\{z \in \mathbb{R} : P(x, z) = k\, Q(x, z)\},$$

for all $x \in \mathcal{L}_k$, where

$$\mathcal{L}_k = \{x > 0 : P(x, \cdot) = k\, Q(x, \cdot) \text{ admits a solution.}\}$$

Moreover, for convenience, we include the case $k = \infty$, where the minimal and maximal solutions of $Q(x, \cdot) = 0$ (instead of $L(x, \cdot) = k$) are considered; the domain $\mathcal{L}_\infty$ is also defined. One easily checks that

$$\mathbb{T} = \{(x, z) : x \in \mathcal{L}_0,\ z = T_u(x, 0) \text{ or } T_d(x, 0)\}, \text{ and}$$

$$\mathbb{B} = \{(x, z) : x \in \mathcal{L}_\infty,\ z = T_u(x, \infty) \text{ or } T_d(x, \infty)\}.$$

Functions T_u and T_d allow us to define a subclass of solutions to $g' = L(\cdot, g)$:

Definition 6.3. A continuous solution $g : \mathcal{D} \to \mathbb{R}$ is said to be a **maximal inner solution** if

(1) $T_d(x, 0) \le g(x) \le T_u(x, 0)$, for all $x \in \mathcal{D}$, and
(2) g cannot be extended to an interval strictly larger than $\mathcal{D}$, without violating either (1) or the continuous-solution property.

FIGURE 7.

Thanks to the local Lipschitz property of the function L away from $\mathbb{B}$, the general theory of ordinary differential equations, namely the Peano Existence Theorem (see, for example Theorem I, p. 73 in [Wal98]), states that, starting from any point (x, z) with $z \in [T_d(\cdot, 0), T_u(\cdot, 0)]$ and $L(x, y)$ well defined, one can construct a maximal inner solution $g : \mathcal{D} \to \mathbb{R}$. This solution is necessarily real-analytic away from the curve $\mathcal{B}$.
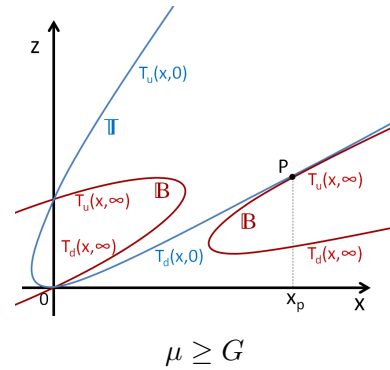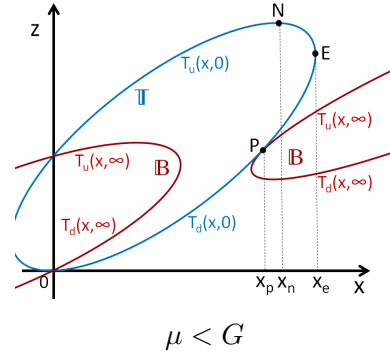
We will be particularly interested in maximal inner solutions started at the top portion of $\mathbb{T}$, i.e., at the point $(\alpha, T_u(\alpha, 0))$, for $\alpha \in \mathcal{L}_0 \setminus \{x_P\}$. Not assuming uniqueness, we pick one such solution, denote it by g_α, its domain by $\mathcal{D}_\alpha$, and the right boundary point of $\mathcal{D}_\alpha$ by $\beta_\alpha \in (0, \infty]$. To avoid the analysis of unnecessary cases, we assume from the start that $\alpha \in (0, x_N) \setminus \{x_P\}$, so that $T_u(\cdot, 0)$ is strictly increasing in the neighborhood of α and the singularity x_P is not used as the initial value (a curious reader can peak ahead to Proposition 6.9, to see how the case $\alpha = x_P$ can be handled.)

To rule out the possible encounters of a maximal inner solution with $\mathbb{B}$ away from the singular point x_P, we delve a bit deeper into the geometry of the right-hand side L of our ODE. We start by a technical lemma which will help us construct the containment curve τ. Some more explicit expressions for the upper and lower curves T_u and T_d are going to be needed:

$$T_{u,d}(x,k) = \frac{1-p}{2\delta} + \frac{b(k)x \pm \sqrt{(b(k)^2 - 4a(k)c(k))x^2 + 4p(1-p)(k+1)(b(k)-4\delta k)x + 4p^2(1-p)^2(1+k)^2}}{2a(k)} \qquad (6.0)$$

where $a(k), b(k), c(k)$ are given by

$$\begin{cases} a(k) = 2p\delta(1+k), \\ b(k) = (2\delta + p(1-p)(2\mu - \sigma^2))k + 2p(1-p)\mu, \\ c(k) = (1-p)(2\mu + (p^2-1)\sigma^2)k + (1-p)^3\sigma^2. \end{cases} \qquad (6.1)$$

The end-points of the domains of T_u and T_d, i.e., those x for which $T_u(x,k) = T_d(x,k)$ are given by

$$x_\pm(k) = \frac{k+1}{\pm G(1 - \frac{p}{1-p}k) - \mu + (\frac{A}{1-p} - \mu)k}.$$

We can also check that $T_{u,d}(x, \infty) := \lim_{k \to \infty} T_{u,d}(x,k)$ are solutions of $Q(x, \cdot) = 0$ and that $x_\pm(\infty) := \frac{1}{\frac{p}{1-p}(A \mp G) + (A - \mu)}$ are the solutions to $T_d(\cdot, \infty) = T_u(\cdot, \infty)$. Finally, we note for future reference that $0 < x_-(\infty) < x_+(\infty) \leq x_P$ holds, and that, for $\mu < G$, the x-coordinates of the north and east points (N, E) are given by $x_N = \frac{2\mu}{G^2 - \mu^2}$ and $x_E = x_+(0) = \frac{1}{G-\mu}$.

Lemma 6.4. *For $0 < p < 1$ and $\mu < A$, there exists constant $k_0 \in (0, \frac{1-p}{p})$ such that*

(1) $x_+(k_0) \geq x_P$ and $(0, x_+(k_0)] \subseteq \mathcal{L}_{k_0}$.
(2) $\frac{\partial}{\partial x}T_u(x, k_0) < k_0$ for $x \in (0, x_+(k_0))$.

Proof. (1) A direct calculation shows that, for $k = \frac{1-p}{p}$, we have $x_+(k) = x_P > 0$, as well as

$$b^2(k) - 4a(k)c(k) = 4(1-p)^2(A-\mu)^2 > 0, \text{ and } k - \frac{d}{dx}T_u(0, k) = \frac{1}{\delta}(1-p)(A-\mu) > 0.$$

By continuity, we can find $k_0 \in (0, \frac{1-p}{p})$ such that

$$b(k_0)^2 - 4a(k_0)c(k_0) > 0, \ k_0 > \frac{\partial}{\partial x}T_u(0, k_0) \text{ and } x_+(k_0) > 0.$$

We can check that $x_+(k_0) \leq x_-(k_0)$, which, in turn, implies that $\mathcal{L}_{k_0} = (0, x_+(k_0)] \cup [x_-(k_0), \infty)$. Since $x_+(\frac{1-p}{p}) = x_P$ and $\frac{d}{dk}x_+(\frac{1-p}{p}) \leq 0$, we conclude that $x_+(k_0) \geq x_P$

(2) The result follows from $\frac{\partial^2}{\partial x^2}T_u(x, k_0) < 0$ and $\frac{\partial}{\partial x}T_u(0, k_0) < k_0$. $\qquad \square$

With the constant k_0 as in Lemma 6.4 above fixed, we define the **containment curve** τ : $[0, \infty) \to \mathbb{R}$ and a **containment region** $\Omega_0 \in \mathbb{R}^2$ by

$$\tau(x) = \max_{x' \in [0,x]} T_u(x' \wedge x_+(k_0), k_0), \text{ and } \Omega_0 = \{(x, z) \in (0, \infty) \times \mathbb{R} : T_d(x, 0) \vee \tau(x) \leq z \leq T_u(x, 0)\}.$$

The significance of these objects is made clearer in the following proposition. The reader is invited to consult Figure 2 for an illustration.

Proposition 6.5. *For $0 < p < 1$ and $\mu < A$, the following statements hold:*

(1) If $\pi \geq 1$, then $\Omega_0 \cap \mathbb{B} = \{P\}$, $\tau(x_P) = T_u(x_P, 0)$, and $x_P \leq x_N$. For $\pi < 1$, $\Omega_0 \cap \mathbb{B} = \emptyset$.

(2) Ω_0 is simply connected. It is bounded if and only if $\mu < G$.

(3) $(\Omega_0 \setminus \{P\}) \cap \{L(x,z) = k\} = (\Omega_0 \setminus \{P\}) \cap \{z = T_u(x,k)$ or $z = T_d(x,k)\}$.

(4) $(\Omega_0 \setminus \{P\}) \cap \{L(x,z) > k\} = (\Omega_0 \setminus \{P\}) \cap \{T_d(x,k) < z < T_u(x,k)\}$.

(5) $0 \leq L(x,z) \leq k_0$ for $(x,z) \in \Omega_0 \setminus \{P\}$.

(6) $\tau \in C^1([0,\infty))$. For $x > 0$ such that $T_d(x,0) < \tau(x) < T_u(x,0)$, we have $\tau'(x) < L(x, \tau(x))$.

Proof. (1) It is easily checked that, when all sets are seen as subsets of $(0,\infty) \times \mathbb{R}$, that

$$\Omega_0 \cap \mathbb{B} = \Big\{T_d(x,0) \vee \tau(x) \leq z \leq T_u(x,0)\Big\} \bigcap \Big(\Big\{x \leq x_-(\infty),\ z = T_d(x,\infty) \text{ or } T_u(x,\infty)\Big\} \cup \{P\}\Big).$$

Hence, it will be enough to show that following two claims hold:

Claim 1: *For $0 < x \leq x_-(\infty)$, we have $\tau(x) > T_u(x,\infty)$.* Since $T_u(\cdot, \frac{1-p}{p})$ is a straight line on $[0, x_P]$ and $T_u(\cdot, \infty)$ is concave on $[0, x_-(\infty)]$, the easy-to-check facts that

$$T_u(0, \tfrac{1-p}{p}) = T_u(0, \infty) \text{ and } \tfrac{\partial}{\partial x} T_u(0, \tfrac{1-p}{p}) > T_u'(0, \infty)$$

imply that $T_u(\cdot, \frac{1-p}{p}) > T_u(\cdot, \infty)$, on $(0, x_-(\infty)]$. Similarly, $T_u(\cdot, k_0) > T_u(\cdot, \frac{1-p}{p})$ on $(0, x_-(\infty)]$.

Claim 2: *$\tau(x_P) \geq z_P$, with equality if and only if $\pi \geq 1$.* Several sub-cases are considered:

i) $\pi \geq 1$: Then, $\frac{\delta}{p} > \frac{(1-p)\sigma^2}{2}$ and $z_P = T_u(x_P, k_0)$. Since $T_u(\cdot, k_0)$ is strictly concave and $\frac{\partial}{\partial x} T_u(x_P, k_0) = \frac{2(1-p)^2 \sigma^2 (\pi - 1)}{2\delta - p(1-p)\sigma^2} \geq 0$, the map $T_u(\cdot, k_0)$ is strictly increasing on $[0, x_P]$. Thus, $\tau(x_P) = T_u(x_P, k_0) = z_P$.

ii) $\pi < 1$, $\frac{\delta}{p} > \frac{(1-p)\sigma^2}{2}$: $\frac{\partial}{\partial x} T_u(x_P, k_0) < 0$ implies that $\tau(x_P) > T_u(x_P, k_0) = z_P$.

iii) $\pi < 1$, $\frac{\delta}{p} < \frac{(1-p)\sigma^2}{2}$: $\tau(x_P) \geq T_u(x_P, k_0) > T_d(x_P, k_0) = z_P$.

iv) $\pi < 1$, $\frac{\delta}{p} = \frac{(1-p)\sigma^2}{2}$: $\lim_{x \nearrow x_P} \frac{\partial}{\partial x} T_u(x, k_0) = -\infty$ implies $\tau(x_P) > T_u(x_P, k_0) = z_P$.

(2) For the simple connectedness of Ω_0, it is enough to show that $\{x > 0 : T_d(x,0) \vee \tau(x) \leq T_u(x,0)\}$ is an interval. Given that $T_d(x,0) \leq T_u(x,0)$, for all x, it is enough to show that $\{x > 0 : \tau(x) \leq T_u(x,0)\}$ is an interval. With $x_m \in \text{argmax}_{x \in [0, x_+(k_0)]} T_u(x, k_0)$, similarly as in the proof of Claim 1, we observe that for $x \in [0, x_m]$, $T_u(x, k_0) \leq T_u(x, 0)$ and that $T_u(\cdot, 0)$ strictly increases. Therefore, $(0, x_m] \subset \{x > 0 : \tau(x) \leq T_u(x, 0)\}$. Since $T_u(\cdot, 0)$ is strictly concave and τ is constant after x_m, we have $T_u(\cdot, 0) < \tau$, to the right of the right-most point at which $T_u(\cdot, 0)$ equals τ.

Since $T_u(\cdot, 0)$ is strictly concave and τ is constant after x_m, boundedness of Ω_0 is equivalent to the boundedness of the domain $\mathcal{L}_0$, of $T_u(\cdot, 0)$ and $T_d(\cdot, 0)$. The set $\mathcal{L}_0$ is, in turn, bounded, if and only if $\mu < G$.

(3) The statement follows from definitions of T_u and T_d, and the fact that P is the unique singular point.

(4) We only need to observe that $(\Omega_0 \setminus \{P\}) \subset \{P(x,z) \geq 0\} \cap \{Q(x,z) > 0\}$ from (1).

(5) The result follows from (3),(4) and the definition of τ.

(6) C^1-smoothness of τ follows easily from the construction. With x_m as in (2) above, we have

$$\tau(x) = \begin{cases} T_u(x, k_0), & \text{for } x \in [0, x_m], \\ T_u(x_m, k_0), & \text{for } x \in [x_m, \infty). \end{cases}$$

For $x \in [0, x_m)$, by Lemma 6.4, we have

$$\tau'(x) = \frac{\partial}{\partial x} T_u(x, k_0) < k_0 = L(x, T_u(x, k_0)) = L(x, \tau(x)).$$

For $x \in [x_m, \infty)$, $\tau'(x) = 0$, but the condition $T_d(x, 0) < \tau(x) < T_u(x, 0)$ implies that

$$(x, \tau(x)) \in \{Q(x, z) > 0, \ P(x, z) > 0\} \subset \{L(x, z) > 0\}. \quad \square$$

With the result of Proposition 6.5 in hand, we can say more about the shape of the function g_α and its domain $\mathcal{D}_\alpha$. In particular, we show that the graph Γ_α stays at a positive distance from any point of $\mathbb{B}$, except, maybe, P. Remember that $x_N = \infty$ if $\mu \geq G$.

Proposition 6.6. *For $\alpha \in (0, x_N)$ such that $(\alpha, T_u(\alpha, 0)) \neq P$, we have*

(1) $(\Gamma_\alpha \setminus \{P\}) \cap \{(x, T_u(x, 0)) : \alpha < x \leq x_N\} = \emptyset$.

(2) $\Gamma_\alpha \subseteq \Omega_0$.

(3) $\mathcal{D}_\alpha$ is a closed interval of the form $[\alpha, \beta_\alpha]$, for some $\beta_\alpha \in (\alpha, \infty]$. If $\beta_\alpha < \infty$ and $(\beta_\alpha, g_\alpha(\beta_\alpha)) \neq P$, then $g_\alpha'(\beta_\alpha) = 0$.

Proof. (1) Suppose that it is not true. Then, there exists $x_0 \in (\alpha, x_N]$ such that $g_\alpha(x_0) = T_u(x_0, 0)$ and $(x_0, g_\alpha(x_0)) \neq P$. Since $T_u(x, 0) \geq g_\alpha(x)$ for $x < x_0$ close enough, we have $\frac{\partial}{\partial x} T_u(x_0, 0) \leq g_\alpha(x_0)$. Combine this with $g_\alpha'(x_0) = 0$, we deduce that $x_0 = x_N$ and $\mu < G$. So, $g_\alpha(x_N) = T_u(x_N, 0)$ and $g_\alpha'(x_N) = 0 = \frac{\partial}{\partial x} T_u(x_N, 0)$. Using this, we can calculate $g_\alpha''(x_N) = \frac{d}{dx} L(x, g_\alpha(x))|_{x=x_N} = 0 > \frac{\partial^2}{\partial x^2} T_u(x_N, 0)$, which contradict to the fact that $T_u(x) \geq g_\alpha(x)$ for $x < x_N$ close enough.

(2) Noting that $(\alpha, g_\alpha(\alpha)) \in \Omega_0$, we assume that there exists a point $x \in \mathcal{D}_\alpha$ with $(x, g_\alpha(x)) \notin \Omega_0$. With x_0 denoting the infimum of all such points, we observe immediately that $x_0 > \alpha$ and $x_0 < x_E \in (0, \infty]$. If, additionally, $x_0 \neq x_P$, we can use the continuity of g_α to conclude that $g_\alpha(x_0) = \tau(x_0)$ and $g_\alpha'(x_0) \leq \tau'(x_0)$. By using (1), we can exclude the case $g_\alpha(x_0) \in \{T_u(x_0, 0), T_d(x_0, 0)\}$, otherwise, D_α should be $[\alpha, x_0]$, which contradicts to the choice of x_0. Then, since $T_d(x_0, 0) < \tau(x_0) < T_u(x_0, 0)$, we reach a contradiction with part (6) of Lemma 6.5:

$$L(x_0, g_\alpha(x_0)) = g_\alpha'(x_0) \leq \tau'(x_0) < L(x_0, \tau(x_0)) = L(x_0, g_\alpha(x_0)).$$

In the case $x_0 = x_P$, we have $g_\alpha(x_0) = \tau(x_0)$ and, by the definition of the point x_0 and the domain Ω_0, there exists $x' > x_0$, $x' \in \mathcal{D}_\alpha$ such that $g_\alpha(x') < \tau(x')$. Consequently, we have $g_\alpha'(x'') < \tau'(x'')$ for some $x_0 < x'' < x'$, and we observe that $L(x'', g_\alpha(x'')) \geq L(x'', \tau(x''))$, because $L(x, z)$ is a decreasing function of z near P. Now we reach a contradiction as in the case $x_0 \neq x_P$:

$$L(x'', g_\alpha(x'')) = g_\alpha'(x'') < \tau'(x'') \leq L(x'', \tau(x'')) \leq L(x'', g_\alpha(x'')),$$

where $\tau'(x'') \leq L(x'', \tau(x''))$ can be showed by part (6) of Lemma 6.5.

(3) We first observe that the initial value $(\alpha, g_\alpha(\alpha))$ lies on the graph of the function $T_u(\cdot, 0)$ which is strictly increasing in the neighborhood of α. Therefore, any extension of g_α to the left of α would cross $\mathbb{T}$ and exit the set $\{z \leq T_u(x, 0)\}$. So, left end-point of the domain $\mathcal{D}_\alpha$ should be α.

To deal with the right end-point of $\mathcal{D}_\alpha$, we note that noe of the following must occur: 1) g_α explodes, 2) $\mathbb{B}$ cossed, or 3) $\mathbb{T}$ is hit. The first possibility is easily ruled out by the observation that no explosion can happen without g_α crossing the curve $\mathbb{T}$, first. The second possibility is severely limited by (2) above; indeed, with part (1) of Proposition 6.5, $\Gamma_\alpha \cap \mathbb{B} \subset \{P\}$. It is clear now that, in the right end-point limit β_α, the function g_α hits $\mathbb{T}$, provided $\beta_\alpha < \infty$. For $\beta_\alpha < \infty$ and $(\beta_\alpha, g_\alpha(\beta_\alpha)) \neq P$, $\lim_{x \nearrow \beta_\alpha} g_\alpha(x)$ clearly exists, and, so, $\beta_\alpha \in \mathcal{D}_\alpha$. Furthermore, $g_\alpha'(\beta_\alpha) = 0$ since $(\beta_\alpha, g_\alpha(\beta_\alpha)) \in \mathbb{T} \setminus \{P\}$. In case $\beta_\alpha < \infty$ and $(\beta_\alpha, g_\alpha(\beta_\alpha)) = P$, we also conclude that $\beta_\alpha \in \mathcal{D}_\alpha$, by observing that $\tau(x) \leq g_\alpha(x) \leq T_u(x, 0)$ for $x < x_P$ and $\lim_{x \nearrow x_P} T_u(x, 0) = \lim_{x \nearrow x_P} \tau(x)$. $\square$

6.2. **The sub-case** $\mu < G$. We focus on the case $\mu < G$ in this subsection. The curve $\mathbb{T}$ is now an ellipse and it admits a north pole with the x-coordinate $x_N < \infty$. By part (1) of Proposition 6.5 and (2) of Proposition 6.6, we have the following dichotomy, valid for all $\alpha \in (0, x_N)$.

(1) For $\pi < 1$, $P \notin \Gamma_\alpha$, and
(2) For $\pi \geq 1$, $P \in \Gamma_\alpha$ if and only if $x_P \in \mathcal{D}_\alpha$.

We start with the first possibility which avoids the singularity P altogether.

Proposition 6.7 $(0 < p < 1, \mu < G, \pi < 1)$. *Suppose that $0 < p < 1$, $\mu < G$ and $\pi < 1$. Then, $\beta_\alpha \in (\alpha, x_E]$ and g_α is of class C^∞ on $\text{Int}\,\mathcal{D}_\alpha$, for all $\alpha \in (0, x_N)$. Moreover, the function $G(\alpha) = \int_\alpha^{\beta_\alpha} \frac{g_\alpha{}'(x)}{x}\,dx$ has the following properties:*

$$G \text{ is continuous on } (0, x_N), \lim_{\alpha \searrow 0} G(\alpha) = +\infty, \text{ and } \lim_{\alpha \nearrow x_N} G(\alpha) = 0. \qquad (6.2)$$

In particular, g_α solves the free-boundary problem (2.10), (2.12), for some $\alpha \in (0, x_N)$.

Proof. By part (2) of Proposition 6.5, β_α is bounded and in $(\alpha, x_E]$. $g_\alpha{}'(\alpha) = g_\alpha{}'(\beta_\alpha) = 0$ is a consequence of part (3) of Proposition 6.6. Since $P \notin \Gamma_\alpha$, smoothness of g_α follows from the general theory (Peano's theorem). Moreover, the existence of the initial value α, with the desired properties, is a direct consequence of the listed properties of G, by way of the intermediate value theorem. We, therefore, focus on (6.2) in the remainder of the proof, which is broken into several claims. The proof of each claim is placed directly after the corresponding statement.

Claim 1: *If $g_\alpha(\beta_\alpha) = T_u(\beta_\alpha, 0)$, then $\beta_\alpha > x_N$.* This follows from part (1) of Proposition 6.6.

Claim 2: *The map $\alpha \mapsto \beta_\alpha$ is continuous.* For this, we use the implicit-function theorem and the continuity of g_α with respect to the initial data (see, e.g., Theorem VI., p 145 in [Wal98]). To be able to use the implicit-function theorem, it will be enough to observe that, $g_\alpha(\cdot)$ is not tangent to $T_u(\cdot, 0)$ (or $T_d(\cdot, 0)$) at $x = \beta_\alpha$, which is a consequence of $g_\alpha{}'(\beta_\alpha) = 0$ and Claim 1. above.

Claim 3: *The map $\alpha \mapsto G(\alpha)$ is continuous.* It suffices to use the dominated convergence theorem. Its conditions are met, since $g_\alpha{}'(x) \in [0, k_0]$ (by Proposition 6.5, part (5)).

Claim 4: $\lim_{\alpha \searrow 0} G(\alpha) = \infty$. The joint continuity of $\frac{\partial}{\partial x} T_u(x, k)$ at $(0, 0)$ and the fact that $\frac{\partial}{\partial x} T_u(0, 0) = \frac{(1-p)\mu}{\delta} > 0$, imply that there exists $\epsilon > 0$ such that

$$\frac{\partial}{\partial x} T_u(x, \epsilon) > 2\epsilon \text{ for } x \in [0, \epsilon].$$

We define $l(\alpha) = \alpha + \frac{T_u(\alpha, 0) - T_u(\alpha, \epsilon)}{\epsilon}$ and remind the reader that $T_u(0, k) = \frac{1-p}{\delta}$ for each k, so that $\lim_{\alpha \searrow 0} l(\alpha) = 0$. Hence, we can pick $\alpha_\varepsilon > 0$ such that $l(\alpha) < \varepsilon$, for $\alpha < \alpha_\varepsilon$.

For any given $\alpha \in (0, \alpha_\epsilon)$, if it so happens that $g_\alpha(x) > T_u(x, \epsilon)$ for $x \in [\alpha, l(\alpha)]$, then Proposition 6.5, part (4), implies that $g_\alpha{}'(x) < \epsilon$ on $[\alpha, l(\alpha)]$. Therefore,

$$0 < g_\alpha(l(\alpha)) - T_u(l(\alpha), \epsilon) = \int_\alpha^{l(\alpha)} \left(g_\alpha{}'(x) - \frac{\partial}{\partial x} T_u(x, \epsilon)\right)dx + T_u(\alpha, 0) - T_u(\alpha, \epsilon)$$

$$\leq \int_\alpha^{l(\alpha)} (\epsilon - 2\epsilon)dx + T_u(\alpha, 0) - T_u(\alpha, \epsilon) = 0,$$

which is contradiction. We conclude that g_α intersects $T_u(\cdot, \epsilon)$ on $[\alpha, l(\alpha)]$, for each $\alpha \in (0, \alpha_\varepsilon)$.

Using the fact that $\frac{\partial}{\partial x} T_u(x, \epsilon) > L(x, T_u(x, \epsilon))$ on $[0, \epsilon]$, we conclude that $g_\alpha(x) < T_u(x, \epsilon)$ on $[l(\alpha), \epsilon]$. By Proposition 6.5, part (4) and the fact that $\tau(x) > T_d(x, \epsilon)$ for small x, we have that

$g_\alpha{}'(x) \geq \epsilon$ on $[l(\alpha), \epsilon]$. Therefore,

$$\liminf_{\alpha \searrow 0} G(\alpha) \geq \liminf_{\alpha \searrow 0} \int_{l(\alpha)}^{\epsilon} \frac{\epsilon}{x} dx = \liminf_{\alpha \searrow 0} \epsilon \ln \left(\frac{\epsilon}{l(\alpha)} \right) = \infty.$$

Claim 5: $\lim_{\alpha \nearrow x_N} G(\alpha) = 0$. We start with the inequality $T_u(\alpha, 0) = g_\alpha(\alpha) < g_\alpha(\beta_\alpha) \leq T_u(x_N, 0)$, which implies that $\lim_{\alpha \nearrow x_N} \beta_\alpha = x_N$. Thus, by Proposition 6.5, part (5), we have

$$\limsup_{\alpha \nearrow x_N} G(\alpha) \leq \limsup_{\alpha \nearrow x_N} \frac{(\beta_\alpha - \alpha)k_0}{\alpha} = 0. \qquad \square$$

Before we move on to the case $\pi \geq 1$, we need a few facts about a specific, singular, ODE.

Lemma 6.8. *Given $\varepsilon > 0$, consider the ODE*

$$h'(y) = -\frac{h(y)}{A(y)y^2} + B(y), \tag{6.3}$$

where $A, B : [-\varepsilon, \varepsilon]$ are continuous functions, with $A(0) > 0$. Then, the following statements hold:
(1) There is a single solution h_+ of (6.3) on $(0, \varepsilon]$ with $\lim_{y \searrow 0} h_+(0) = 0$.
(2) No solutions h_+ exist with $\lim_{y \searrow 0} h_+(y) = c \in \mathbb{R} \setminus \{0\}$.
(3) For any solution h_- on $[-\varepsilon, 0)$, we have $\lim_{y \nearrow 0} h_-(y) = 0$.
(4) Any function $h : [-\varepsilon, \varepsilon] \to \mathbb{R}$ of the form

$$h(y) = h_+(y)\mathbf{1}_{y>0} + h_-(y)\mathbf{1}_{y<0},$$

where h_+ is as in (1) above, and h_- is any *function as in (3) above, is a C^1-solution to (6.3).*

Proof. Elementary transformations can be used to show that for any solution h of (6.3) defined on $[-\varepsilon, \varepsilon] \setminus \{0\}$, there exist constants c_1 and c_2 such that

$$h(y) = \begin{cases} e^{D(y)}\left(c_1 - \int_y^\varepsilon B(t)e^{-D(t)}dt\right), & y \in (0, \varepsilon], \\ e^{D(y)}\left(c_2 + \int_{-\varepsilon}^y B(t)e^{-D(t)}dt\right), & y \in [-\varepsilon, 0), \end{cases} \quad \text{where } D(y) = \begin{cases} \int_y^\varepsilon \frac{1}{A(t)}\frac{dt}{t^2}, & y \in (0, \varepsilon], \\ \int_{-\varepsilon}^y -\frac{1}{A(t)}\frac{dt}{t^2}, & y \in [-\varepsilon, 0) \end{cases}$$

We first note that $\lim_{y \searrow 0} D(y) = \infty$. So, to satisfy the condition $\lim_{y \searrow 0} h_+(y) = c \in \mathbb{R}$, the only possibility is $c_1 = \int_0^\varepsilon B(t)e^{-D(t)}dt$. Then, the L'Hospital's rule implies that

$$\lim_{y \searrow 0} h_+(y) = \lim_{y \searrow 0} e^{D(y)} \int_0^y e^{-D(t)}B(t)\,dt = \lim_{y \searrow 0} \frac{B(y)}{\frac{1}{A(y)y^2}} = 0,$$

and we immediately conclude (1) and (2).

As far as (3) is concerned, since $\lim_{y \nearrow 0} e^{D(y)}c_2 = 0$, for any c_2, the limiting behavior is independent of c_2. Moreover, another use of the L'Hospital's rule implies that $h(y) \to 0$, as $y \nearrow 0$, for each $c_2 \in \mathbb{R}$.

It remains to show (4), and, for this, we start by computing the derivative at 0 of h. Like above, we use the L'Hospital rule and the explicit expression for h:

$$\lim_{y \searrow 0} \frac{h(y) - h(0)}{y} = \lim_{y \searrow 0} \frac{B(y)}{1 + \frac{1}{A(y)y}} = 0.$$

Similarly, $\lim_{y \nearrow 0} \frac{h(y) - h(0)}{y} = 0$, and, so $h'(0) = 0$. To establish that $\lim_{y \to 0} h'(y) = h'(0) = 0$, we first use the L'Hospital rule to compute $\lim_{y \to 0} \frac{h(y)}{y^2} = \frac{B(0)}{1/A(0)}$, so that, using the equation (6.3) for h, we can immediately deduce that $\lim_{y \to 0} h'(y) = 0$. $\qquad \square$

Proposition 6.9 ($0 < p < 1$, $\mu < G$, $\pi \geq 1$)**.** *Suppose that $0 < p < 1$, $\mu < G$.*
(1) If $\pi > 1$ and

a) $\alpha \in (x_P, x_N)$. *Then* $\beta_\alpha \in (\alpha, x_E]$, g_α *is of class* C^∞ *and* $P \notin \Gamma_\alpha$.

b) $\alpha = x_P$. *Then the limits*

$$\beta_{x_P} = \lim_{\alpha \searrow x_P} \beta_\alpha, \ \ and \ \ g_{x_P}(x) = \lim_{\alpha \searrow x_P} g_\alpha(x), \ \ x \in (x_P, \beta_{x_P}],$$

exist and define a continuous solution to (2.10) *with the domain* $[x_P, \beta_{x_P}]$.

c) $\alpha \in (0, x_P)$. *Then* $\beta_\alpha \in (\alpha, x_E]$, g_α *is of class* C^2, $P \in \Gamma_\alpha$ *and* $g_\alpha'(x_P) = \frac{\partial}{\partial x} T_u(x, 0)$.

(2) *If* $\pi = 1$, *then* $x_N = x_P$. *For* $\alpha \in (0, x_N)$, $(\beta_\alpha, g_\alpha(\beta_\alpha)) = P$ *and* g_α *is of class* C^2.

In all these cases, the function $G(\alpha) = \int_\alpha^{\beta_\alpha} \frac{g_\alpha'(x)}{x} \, dx$ *has the following properties:*

$$G \text{ is continuous on } (0, x_N), \ \lim_{\alpha \searrow 0} G(\alpha) = +\infty, \ \lim_{\alpha \nearrow x_N} G(\alpha) = 0. \tag{6.4}$$

In particular, g_α *is a solution to the free-boundary problem* (2.10), (2.12), *for some* $\alpha \in (0, x_N)$.

Remark 6.10. (1) The parameter regime treated in Proposition 6.9 above leads to a truly singular behavior in the ODE (2.10). Indeed, the maximal continuous solution passes through the singular point P, at which the right-hand side $L(\cdot, g)$ is not well-defined. It turns out that the continuity of the solution, coupled with the particular form (2.10) of the equation, forces higher regularity (we push the proof up to C^2) on the solution. The related equation (6.3) of Lemma 6.8 provides a very good model for the situation. Therein, uniqueness fails on one side of the equation (and general existence on the other), but the equation itself forces a smooth passage of any solution through the origin. It follows immediately, that, even though high regularity can be achieved at the singularity, the solution will never be real analytic there, except, maybe, for one particular value of $\log(\frac{1+\bar\lambda}{1-\underline\lambda})$. This is a general feature of singular ODE with a rational right-hand sides. Consider, for example, the simplest case $y' = -\frac{y}{x^2}$ which admits as a solution the textbook example $y(x) = e^{1/x} \mathbf{1}_{\{x<0\}}$ of a C^∞ function which is not real analytic.

(2) For large-enough $\log(\frac{1+\bar\lambda}{1-\underline\lambda})$, the value of α such that g_α solves (2.10), (2.12), will fall below x_P, and an interesting phenomenon will occur. Namely, the right free boundary $\bar x$ will stop depending on $\bar\lambda$ or $\underline\lambda$. Indeed, the passage through the singularity P simply "erases" the memory of the initial condition in g_α. In financial terms, the right boundary of the no-trade region will be stop depending on the transaction costs, while the left boundary will continue to open up as the transaction costs increase.

Proof. We will only prove (1) here; (2) can be proved by the same methods used in the proof of c) below. For both (1) and (2), $g_\alpha'(\alpha) = g_\alpha(\beta_\alpha) = 0$ follows easily.

a) By Proposition 6.5, part (1), $\Omega_0 \cap \mathbb{B} \cap \{(x, z) : x > x_P\} = \emptyset$. So, if $\alpha > x_P$, the statement can be proved by using the argument from the proof of Proposition 6.7, mutatis mutandis.

b) The existence of the limit β_{x_P} from the statement is established in a matter similar to that used to prove the continuity of the map $\alpha \to \beta_\alpha$ in Claim 2. in the proof of Proposition 6.7. The existence of the limit g_{x_P} follows from a standard argument involving a weak formulation and the dominated convergence theorem. Finally, by part (2) of Proposition 6.6 and $T_u(x_P, 0) = \tau(x_P)$, we conclude that g_{x_P} is defined and continuous on $[x_P, \beta_{x_P}]$, with $(x_P, g_{x_P}(x_P)) = P$.

c) As in the proof of Proposition 6.7, $g_\alpha(x)$ does not hit either τ or $T_u(\cdot, 0)$ on $(0, x_P)$. Hence, we must have $x_P \in \mathrm{Cl}\,\mathcal{D}_\alpha$; moreover since the curves $T_u(\cdot, 0)$ and $\tau(\cdot)$ coalesce at x_P, the limit $\lim_{x \to x_P} g_\alpha(x)$ exists and equals to $T_u(x_P, 0)$. In particular, we have $x_P \in \mathcal{D}_\alpha$ and $P \in \Gamma_\alpha$.

For $x_p < x_N$, part b) above guarantees that a continuous solution with a domain of the form $[x_P, \beta_{x_P}]$, with $\beta_{x_P} > x_P$, exists. Therefore, by maximality, a maximal inner solution g_α, with $\mathcal{D}_\alpha = [\alpha, \beta_{x_P}]$ exists (in other words, $\beta_\alpha = \beta_{x_P}$, for all $\alpha < x_P$).

Our next task is to upgrade the regularity of g_α from $C[\alpha, \beta_\alpha]$ to $C^2[\alpha, \beta_\alpha]$, where, clearly, we can focus on a neighborhood of the point x_P: we need to show that that $g_\alpha'(x_P), g_\alpha''(x_P)$ exist and $g_\alpha'(x), g_\alpha''(x)$ are continuous at x_P. The argument is divided in several claims, whose proofs follow the respective statements.

Claim 1: $g_\alpha'(x)$ *does not admit a local minimum on* $(\alpha, x_P) \cup (x_P, \beta_\alpha)$. Suppose, to the contrary, that it does. Then, there exists $\varepsilon > 0$ and a point $x_m \in (\alpha, x_P) \cup (x_P, \beta_\alpha)$ such that

$$g_\alpha'(x_m) \le g_\alpha'(x) \text{ for } x \in [x_m - \epsilon, x_m + \epsilon].$$

For $k_m := g_\alpha'(x_m)$, parts (3) and (4) of Proposition 6.5 imply that

$$\begin{cases} g_\alpha(x_m) = T_u(x_m, k_m) \text{ or } g_\alpha(x_m) = T_d(x_m, k_m), \\ T_d(x, k_m) \le g_\alpha(x) \le T_u(x, k_m) \text{ on } [x_m - \epsilon, x_m + \epsilon]. \end{cases} \tag{6.5}$$

We focus on the case $g_\alpha(x_m) = T_u(x_m, k_m)$, with the other one - when $g_\alpha(x_m) = T_d(x_m, k_m)$ - being similar. By (6.5), we have $\frac{\partial}{\partial x} T_u(x_m, k_m) = g_\alpha'(x_m) = k_m$; moreover, since $\frac{\partial^2}{\partial x^2} T_u(x, k) < 0$, we get $0 = g_\alpha'(x_m) - \frac{\partial}{\partial x} T_u(x_m, k_m) < g_\alpha'(x) - \frac{\partial}{\partial x} T_u(x, k_m)$, on $(x_m, x_m + \epsilon]$. This leads to the following contradiction:

$$0 < \int_{x_m}^{x_m + \epsilon} \Big(g_\alpha'(x) - \tfrac{\partial}{\partial x} T_u(x, k_m) \Big) dx = g_\alpha(x_m + \epsilon) - T_u(x_m + \epsilon, k_m) \le 0.$$

Claim 2: $g_\alpha \in C^1([\alpha, \beta_\alpha])$ *and* $g_\alpha'(x)$ *decreases around* x_P. We observe that $\tau(x) < g_\alpha(x) < T_u(x, 0)$ for $x \in (\alpha, x_P) \cup (x_P, \beta_\alpha)$, $\tau(x_P) = g_\alpha(x_P) = T_u(x_P, 0)$, and $\tau'(x_P) = \frac{\partial}{\partial x} T_u(x_P, 0)$, and conclude that g_α is differentiable at x_P and $g_\alpha'(x_P) = \tau'(x_P) = \frac{\partial}{\partial x} T_u(x_P, 0)$. By the Claim 1., $\lim_{x \nearrow x_P} g_\alpha'(x), \lim_{x \searrow x_P} g_\alpha'(x)$ exist. So, using the mean value theorem, we obtain $g_\alpha'(x_P) = \lim_{x \to x_P} g_\alpha'(x)$ and conclude that $g_\alpha \in C^1([\alpha, \beta_\alpha])$.

Given an ε in a small-enough neighborhood of 0, the concavity of $T_u(\cdot, 0)$ implies that

$$g_\alpha(x_P - \epsilon) < T_u(x_P - \epsilon, 0) < T_u(x_P, 0) - \epsilon \tfrac{\partial}{\partial x} T_u(x_P, 0) = g_\alpha(x_P) - \epsilon g_\alpha'(x_P).$$

The mean value theorem can now be used to conclude that there exist x_1, x_2, arbitrarily close to x_P, with $x_1 < x_P < x_2$ such that

$$g_\alpha'(x_1) > g_\alpha'(x_P) > g_\alpha'(x_2).$$

Finally, if we combine the obtained results with those of Claim 1., we can conclude that $g_\alpha'(x)$ decreases near x_P.

Claim 3: *The second derivative of* g_α *exists at* x_P *and*

$$g_\alpha''(x_P) = -\frac{(1-p)^2 \sigma^2 (2\delta - 2p\mu + p(1-p)\sigma^2)^2 (2\delta + 2(1-p)\mu + (p-2)(1-p)\sigma^2)}{p(2\delta - p(1-p)\sigma^2)^3}. \tag{6.6}$$

The proof is based on an explicit computation where the easy-to-check fact that our ODE admits the form

$$g_\alpha'(x) = -\frac{(g_\alpha(x) - T_u(x, 0))(g_\alpha(x) - T_d(x, 0))}{(g_\alpha(x) - T_u(x, \infty))(g_\alpha(x) - T_d(x, \infty))},$$

is used. We begin with the equality

$$\frac{g_\alpha(x) - g_\alpha(x_P) - g_\alpha'(x_P)(x - x_P)}{(x - x_P)^2} = \frac{T_d(x, \infty) - T_d(x_P, \infty) - \frac{\partial}{\partial x} T_d(x_P, \infty)(x - x_P)}{(x - x_P)^2}$$

$$- \frac{\frac{T_d(x,\infty) - T_u(x,0)}{(x - x_P)^2}}{1 + g_\alpha'(x) \frac{g_\alpha(x) - T_u(x,\infty)}{g_\alpha(x) - T_d(x,0)}}.$$

By L'Hospital's rule, as $x \to x_P$, the right-hand side above converges to

$$\frac{1}{2} \frac{\partial^2}{\partial x^2} T_d(x_P, \infty) - \frac{1}{2} \frac{\frac{\partial^2}{\partial x^2} T_d(x_P, \infty) - \frac{\partial^2}{\partial x^2} T_d(x_P, 0)}{1 + g_\alpha'(x_P) \frac{g_\alpha(x_P) - T_u(x_P,\infty)}{g_\alpha(x_P) - T_d(x_P,0)}},$$

which, in turn, evaluates to the half of the right-hand side of (6.6).

Having computed a second-order quotient of differences for g_α at x_P, we could use the concavity of g_α at x_P (established in Claim 2. above) to conclude that g_α is twice differentiable there. We opt to use a short, self-contained argument, instead, where c denotes the right-hand side of (6.6). For small enough ζ, we have

$$\left(g_\alpha'(x) - g_\alpha'(x_P) - c(x - x_P) \right) \zeta \le g_\alpha(x) - g_\alpha(x - \zeta) - g_\alpha'(x_P)\zeta - c(x - x_P)\zeta$$

$$= -\tfrac{c}{2}\zeta^2 + \left(g_\alpha(x) - g_\alpha(x_P) - g_\alpha'(x_P)(x - x_P) - \tfrac{c}{2}(x - x_P)^2 \right)$$

$$- \left(g_\alpha(x - \zeta) - g_\alpha(x_P) - g_\alpha'(x_P)(x - \zeta - x_P) - \tfrac{c}{2}(x - x_P - \zeta)^2 \right)$$

$$= -\tfrac{c}{2}\zeta^2 + o((x - x_P)^2) + o((x - \zeta - x_P)^2).$$

If we fix $t > 0$ and choose $\zeta = t \, |x - x_P| \, \text{sgn} \left(g_\alpha'(x) - g_\alpha'(x_P) - c(x - x_P) \right)$, we obtain

$$\limsup_{x \to x_P} \left| \frac{g_\alpha'(x) - g_\alpha'(x_P)}{x - x_P} - c \right| \le -\tfrac{c}{2}t,$$

from which the claim follows immediately.

Claim 4: $g_\alpha \in C^2([\alpha, \beta_\alpha])$. For convenience, we change variables as follows

$$y = x - x_P, \quad f(y) = g_\alpha(x) - g_\alpha(x_P) - g_\alpha'(x_P)(x - x_P) - \tfrac{1}{2}g_\alpha''(x_P)(x - x_P)^2.$$

With respect to the new coordinate system, we have $f \in C^1([\alpha - x_P, \beta_\alpha - x_P]) \cap C^2([\alpha - x_P, 0) \cup (0, \beta_\alpha - x_P])$, and $f(0) = f'(0) = f''(0) = 0$; we need to show that $\lim_{y \to 0} f''(y) = f''(0)$. This follows, however, directly from Lemma 6.8, as we obtain the ODE (6.3) if we differentiate the equality $g' = L(\cdot, g)$, and pass to the new coordinates. The coefficient functions $A(y)$ and $B(y)$ admit a rather messy but explicit form which can be used to establish their continuity. Indeed, it turns out that $A(y)$ and $B(y)$ can be represented as continuous transformations of functions of y, $f(y)/y^2$ and $f'(y)/y$, which are, themselves, continuous. Similarly, the condition $A(0) > 0$ imposed in Lemma 6.8 is satisfied because one can use the aforementioned explicit expression to conclude that $A(0) = \lim_{y \to 0} A(y) = \frac{(1-p)\sigma^2(A - \mu)}{2\delta - p(1-p)\sigma^2} > 0$ ☐

6.3. **The sub-case** $G \le \mu < A$. This sub-case is, perhaps the most challenging of all, as it combines the existence of a singularity with a possible failure of the well-posedness of the value function.

For $k \in \mathbb{R}$ let $l_u(k), l_d(k)$ be the (ordered) solutions X_1, X_2 of the quadratic equation $a(k)X^2 - b(k)X + c(k) = 0$, where $a(k)$, $b(k)$ and $c(k)$ are as in (6.1). The analysis in the sequel centers around the constants $C = C(\mu, \sigma, p, \mu)$ and $K = K(\mu, \sigma, p, \mu)$, given by

$$K = \frac{(1-p)(\mu-G)}{(A-\mu)+p(\mu-G)} \text{ and } C = \int_0^K k\Big(\frac{l_u'(k)}{k - l_u(k)} - \frac{l_d'(k)}{k - l_d(k)}\Big)dk. \tag{6.7}$$

Lemma 6.11. *Assume that $0 < p < 1$ and $G \leq \mu < A$. Then*

(1) *K is the smallest solution to $b(\cdot)^2 = 4a(\cdot)c(\cdot)$. Moreover K is nonnegative and $K = 0$ if and only if $\mu = G$.*

(2) *$\Omega_0 \cap \{L(x, z) = k\}$ is bounded if $k > K$ and unbounded otherwise.*

(3) *For $0 \leq k \leq K$, $l_d(k) > k$.*

(4) *For $0 \leq k < K$, we have*

$$\lim_{x \to \infty} \tfrac{\partial}{\partial x}T_{u,d}(x, k) \to l_{u,d}(k) \text{ and } \lim_{x \to \infty} \tfrac{1}{x}\tfrac{\partial}{\partial k}T_{u,d}(x, k) \to l_{u,d}'(k).$$

(5) *There exists a constant $c^* > 0$ such that for $x > c^*$ and $k \in [0, K)$ we have*

$$\Big|\frac{\frac{\partial}{\partial k}T_d(x,k)}{x(k - \frac{\partial}{\partial x}T_d(x,k))}\Big| < c^* + \frac{c^*}{\sqrt{K-k}}, \quad \Big|\frac{\frac{\partial}{\partial k}T_u(x,k)}{x(k - \frac{\partial}{\partial x}T_u(x,k))}\Big| < c^* + \frac{c^*}{\sqrt{K-k}}.$$

(6) *C is well-defined and nonnegative. Moreover, $C = 0$ if and only if $\mu = G$.*

Proof. (1) It follows by direct computation.

(2) It is easily checked that the leading coefficient of $b(k)^2 - 4a(k)c(k)$ (seen as a polynomial in k) is positive. Therefore, $b(k)^2 - 4a(k)c(k) \geq 0$ for $k \in [0, K]$. Since $b(k) - 4\delta k$ is linear in k and its values at $k = 0, K$ are positive, $4p(1-p)(k+1)(b(k) - 4\delta k) > 0$ for $k \in [0, K]$. Thus, the expression inside the square root in (6.0) is positive for $x \geq 0$ and $k \in [0, K]$, which, in turn, implies that for $k \in [0, K]$, $\Omega_0 \cap \{L(x, z) = k\}$ is unbounded.

Similarly, since $b(k)^2 - 4a(k)c(k)|_{k=K+\epsilon} < 0$ for small enough $\epsilon > 0$, we conclude that the domain $\mathcal{L}_{K+\varepsilon}$ of $T_u(\cdot, K + \epsilon)$ is bounded. Part (4) of Proposition 6.5, implies that $\Omega_0 \cap \{L(x, z) > K + \epsilon\}$ is a bounded set for any sufficiently small $\epsilon > 0$. We conclude that $\Omega_0 \cap \{L(x, z) = k\}$ is bounded for $k > K$.

(3) From the definition of $l_d(k)$ we get

$$2a(k)\Big(l_d(k) - k\Big) = b(k) - 4\delta k + 4p\delta k(\tfrac{1-p}{p} - k) - \sqrt{b(k)^2 - 4a(k)c(k)}$$

We already checked that $b(k) - 4\delta k > 0$ for $k \in [0, K]$. Also, $\tfrac{1-p}{p} - k > 0$ for $k \in [0, K]$, since $\tfrac{1-p}{p} - K = \tfrac{1-p}{p} \cdot \tfrac{A-\mu}{A-\mu+p(\mu-G)} > 0$. Thus, $b(k) - 4\delta k + 4p\delta k(\tfrac{1-p}{p} - k) > 0$ for $k \in [0, K]$. Furthermore,

$$\Big(b(k) - 4\delta k + 4p\delta k(\tfrac{1-p}{p} - k)\Big)^2 - \Big(b(k)^2 - 4a(k)c(k)\Big) =$$
$$= 8p^2\delta(1 + k)(\tfrac{1-p}{p} - k)(-2\delta k^2 + p(2\mu - \sigma^2)k + p^2\sigma^2). \tag{6.8}$$

We can now conclude that the left-hand side of (6.8) is positive on $[0, K]$, since the function $k \mapsto (-2\delta k^2 + p(2\mu - \sigma^2)k + p^2\sigma^2)$ is concave and its values at $k = 0, K$ are positive. It follows immediately that $l_d(k) > k$ for $k \in [0, K]$.

(4) This can be shown by the direct computation.

(5) A straightforward (but somewhat tedious) calculation yields that $\tfrac{\partial}{\partial x}T_d(x, k) \to l_d(k)$, as $x \to \infty$, uniformly in $k \in [0, K]$. So, by (3), we can choose c^* such that $\tfrac{\partial}{\partial x}T_d(x, k) - k > \epsilon$ for some $\epsilon > 0$

and all $x > c^*$, $k \in [0, K]$. Also, we can check that there exists a constant c^* such that for $x > 1$ and $k \in [0, K)$ we have

$$\tfrac{1}{x} \tfrac{\partial}{\partial k} T_d(x, k) < c^* + c^* \tfrac{1}{\sqrt{b(k)^2 - 4a(k)c(k)}} < c^* + c^* \tfrac{1}{\sqrt{K-k}},$$

whence the first inequality in the statement of (5) follows. The second one is obtained in a similar manner.

(6) We first observe that $l'_d(k) > 0$ and $l'_u(k) < 0$ for $k \in [0, K)$. Then, the statement follows from the integrability of $1/\sqrt{K - \cdot}$ on $[0, K]$ and the fact that $\left| k \left(\tfrac{l'_u(k)}{k - l_u(k)} - \tfrac{l'_d(k)}{k - l_d(k)} \right) \right| < c^* + c^* \tfrac{1}{\sqrt{K-k}}$, which is, in turn, implied by (4) and (5) above. $\square$

Remark 6.12. In our current parameter range ($0 < p < 1$, $G \le \mu < A$), the level curve $L = 0$ is a hyperbola and the curve $L = k$ is an ellipse for large-enough values of k. In fact, K is the smallest value of $k \ge 0$ such that $L = k$ is a hyperbola (and, therefore, unbounded).

For $G \le \mu < A$, the Merton proportion π cannot take the value 1, so we only consider the cases $\pi < 1$ and $\pi > 1$ in the following proposition:

Proposition 6.13 ($0 < p < 1$, $G \le \mu < A$). *Assuming that $0 < p < 1$ and $G \le \mu < A$, we have the following statements:*

(1) If $\pi < 1$, then $P \notin \Gamma_\alpha$, for each $\alpha > 0$.
(2) If $\pi > 1$ then $P \in \Gamma_\alpha$ if and only if $\alpha \le x_p$.
In both cases, $\beta_\alpha < \infty$. Moreover, for $G(\alpha) = \int_\alpha^{\beta_\alpha} \tfrac{g_\alpha'(x)}{x} \, dx$, we have

$$G \text{ is continuous on } (0, \infty), \ \lim_{\alpha \searrow 0} G(\alpha) = +\infty, \ \text{and} \ \lim_{\alpha \nearrow +\infty} G(\alpha) = C, \tag{6.9}$$

where C is given by (6.7).

Proof. The parts of statements (1) and (2) involving singularities are proved similarly to parallel statements in Proposition 6.9. We show that $\beta_\alpha < \infty$ for $\pi < 1$, with the case $\pi > 1$ being quite similar. Proceeding by contradiction, we suppose that $\beta_\alpha = \infty$, for some $\alpha > 0$. Then, just like in the proof of Proposition 6.9, we can show that $g_\alpha'(x)$ does not admit a local minimum on (α, ∞). Thus, there exists k^* such that $\lim_{x \to \infty} g_\alpha'(x) = k^*$. From Proposition 6.11, part (2), we learn that $k^* \in [0, K]$, whereas from part (3) we conclude that there exists $\epsilon > 0$ such that $l_d(k^* - \epsilon) > k^* + 2\epsilon$. Since $|g_\alpha'(x) - k^*| < \epsilon$ for large enough x, we can use part (4) of Proposition 6.11, to obtain a contradiction

$$\lim_{x \to \infty} \left(\tfrac{\partial}{\partial x} T_d(x, k^* - \epsilon) - g_\alpha'(x) \right) > k^* + 2\epsilon - (k^* + \epsilon) = \epsilon,$$

with the fact that the inequality $g_\alpha'(x) > k^* - \epsilon$ implies that $g_\alpha(x) > T_d(x, k^* - \epsilon)$, for large x.

It remains to prove (6.9). The main idea is to intersect the solution g_α with the (unbounded) level curve $L = K$. If the two points of intersection are denoted by x_u (the intersection is on $T_u(\cdot, K)$) and x_d (intersection on $T_d(\cdot, K)$), with $x_u < x_d$ (see Figure 8), then the integral in (6.9) is split into three integrals on the intervals $[\alpha, x_u]$, $[x_u, x_d]$ and $[x_d, \beta_\alpha]$. The first and the last integrals are then computed using the change of variable $k = g_\alpha'(x)$, while the limit of the middle integral is shown to be zero.
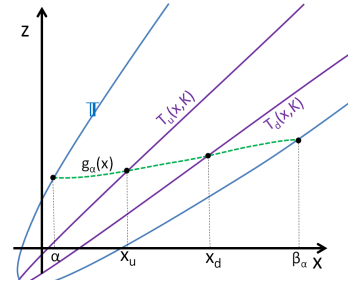

Figure 8. x_u and x_d

We start this program by observing that the region $\Omega_0 \cap \{L(x,z) = K\}$ is unbounded (see Proposition 6.11 (2)), and, hence, so is the region $\Omega_0 \cap \{L(x,z) > K\}$. Also, we observe that $T_u(x,0) > T_u(x,K) > T_d(x,K) > T_d(x,0)$ for $x \in (0,\infty)$. We conclude from there that Γ_α intersects the region

$$\Omega_0 \cap \{(x,z) : T_d(x,K) < z < T_u(x,K)\} = \Omega_0 \cap \{L(x,z) > K\}.$$

Therefore, $g_\alpha'(x_m(\alpha)) > K$ for $x_m(\alpha) \in \operatorname{argmax}_{x \in [\alpha,\beta_\alpha]} g_\alpha'(x)$.

Since $g_\alpha'(x)$ doesn't admit a local minimum on (α, β_α), $x_m(\alpha)$ is uniquely defined and $g_\alpha'(x)$ strictly increases on (α, x_m) and strictly decreases on (x_m, β_α). Consequently, there exists a pair $x_u(\alpha), x_d(\alpha)$ with $x_u(\alpha) \in (\alpha, x_m(\alpha))$ and $x_d(\alpha) \in (x_m(\alpha), \beta_\alpha)$ such that

$$g_\alpha'(x_u(\alpha)) = K, \quad g_\alpha'(x_d(\alpha)) = K.$$

Let $I_\alpha : [0,K] \mapsto [\alpha, x_u(\alpha)]$ be the inverse function of $g_\alpha'(x)$ on $[\alpha, x_u(\alpha)]$, so that

$$g_\alpha'(I_\alpha(k)) = k, \quad g_\alpha(I_\alpha(k)) = T_u(I_\alpha(k), k) \text{ and } I_\alpha'(k) = \frac{\frac{\partial}{\partial k} T_u(I_\alpha(k),k)}{k - \frac{\partial}{\partial x} T_u(I_\alpha(k),k)},$$

where the last equality can be obtained by differentiating the middle one. A change of variables $x = I_\alpha(k)$ yields

$$\int_\alpha^{x_u(\alpha)} \frac{g_\alpha'(x)}{x} dx = \int_0^K \frac{k}{I_\alpha(k)} \frac{\frac{\partial}{\partial k} T_u(I_\alpha(k),k)}{k - \frac{\partial}{\partial x} T_u(I_\alpha(k),k)} dk \xrightarrow{\alpha \to \infty} \int_0^K \frac{k l_u'(k)}{k - l_u(k)} dk, \tag{6.10}$$

where the existence of the limit and its value are obtained using parts (4) and (5) of Proposition 6.11, together with the fact that $\lim_{\alpha \to \infty} I_\alpha(k) = \infty$. In particular, part (5) of Proposition 6.11 allows us to apply the dominated convergence theorem. Similarly, we have

$$\int_{x_d(\alpha)}^{\beta_\alpha} \frac{g_\alpha'(x)}{x} dx \xrightarrow{\alpha \to \infty} - \int_0^K \frac{k l_d'(k)}{k - l_d(k)} dk. \tag{6.11}$$

It remains to show that $\int_{x_u(\alpha)}^{x_d(\alpha)} \frac{g_\alpha'(x)}{x} dt \to 0$ as $\alpha \to \infty$. By Proposition 6.11 (parts (3) and (4)), there exist $\epsilon > 0$ and $x_\epsilon > 0$ such that $\frac{\partial}{\partial x} T_d(x,K) > K + 2\epsilon$, for $x > x_\epsilon$. Moreover, part (2) of the same proposition guarantees the existence of a constant $\alpha_\epsilon > 0$ such that

$$\Omega_0 \cap \{L(x,y) > K + \epsilon\} \subset \{x \le \alpha_\epsilon\}.$$

Then, $g_\alpha'(x) < K + \epsilon$ for $\alpha > \alpha_\epsilon$ and $x \in [\alpha, \beta_\alpha]$, and, so, for $\alpha > \alpha_\epsilon \vee x_\epsilon$, we have

$$\frac{1-p}{\delta} \sqrt{1 + \frac{b(K) - 4\delta K}{p(1-p)(1+K)} x_u(\alpha)} = T_u(x_u(\alpha), K) - T_d(x_u(\alpha), K)$$

$$= g_\alpha(x_u(\alpha)) - T_d(x_u(\alpha), K) + T_d(x_d(\alpha), K) - g_\alpha(x_d(\alpha))$$

$$= \int_{x_u(\alpha)}^{x_d(\alpha)} \left(\frac{\partial}{\partial x} T_d(x,K) - g_\alpha'(x) \right) dx \ge \epsilon (x_d(\alpha) - x_u(\alpha)),$$

where the first equality follows by direct computation, the second one by the fact that $g_\alpha(x_u(\alpha)) = T_u(x_u(\alpha), K)$ and $g_\alpha(x_d(\alpha)) = T_d(x_d(\alpha), K)$, and the final inequality from the choice of α. Hence,

$$\limsup_{\alpha \to \infty} \left| \int_{x_u(\alpha)}^{x_d(\alpha)} \frac{g_\alpha'(x)}{x} dx \right| \le \limsup_{\alpha \to \infty} \left| (K + \epsilon) \ln \left(1 + \frac{x_d(\alpha) - x_u(\alpha)}{x_u(\alpha)} \right) \right|$$

$$\le \limsup_{\alpha \to \infty} \left| (K + \epsilon) \ln \left(1 + \frac{1-p}{\epsilon \delta \, x_u(\alpha)} \sqrt{1 + \frac{b(K) - 4\delta K}{p(1-p)(1+K)} x_u(\alpha)} \right) \right| = 0. \quad \square$$

The remaining task in the proof of Theorem 2.6 is to show that the problem is not well posed, whenever $\log(\frac{1+\bar\lambda}{1-\underline\lambda}) \le C$.

Proposition 6.14. *Assume that $p \in (0,1)$ and $G \le \mu < A$. If $\log(\frac{1+\overline{\lambda}}{1-\underline{\lambda}}) \le C$, where C is defined in (6.7), then $u = \infty$, i.e., the problem is not well posed.*

Proof. Without loss of generality, we assume that $\underline{\lambda} = 0$; indeed, it is enough to scale (the initial value of) the stock price $\{S_t\}_{t \in [0,\infty)}$ by $(1 - \underline{\lambda})$, otherwise.

For $\alpha > 0$, the function g_α in Proposition 6.13 corresponds to the value function u under the transaction costs $\overline{\lambda}$ and $\underline{\lambda} = 0$ such that $G(\alpha) = \log(1 + \overline{\lambda})$, where $G(\alpha) = \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx$. More precisely, Lemma 4.4 in Section 4 above yields that

$$u(1,0,\overline{\lambda},0) = \tfrac{1}{p} g_\alpha(\alpha)^{1-p} = \tfrac{1}{p} T_u(\alpha)^{1-p}, \text{ if } G(\alpha) = \log(1 + \overline{\lambda}), \tag{6.11}$$

where $u(\eta_B, \eta_S, \overline{\lambda}, \underline{\lambda})$ is the optimal utility for the initial position η_B, η_S, under the transaction costs $\overline{\lambda}$ and $\underline{\lambda}$. The strict increase of T_u and the decrease of $u(1,0,\cdot,0)$, imply that $G(\alpha)$ is strictly decreasing, wherever it is defined. It now easily follows that

$$\lim_{\alpha \nearrow \infty} \tfrac{1}{p} g_\alpha(\alpha)^{1-p} = \lim_{\alpha \nearrow \infty} \tfrac{1}{p} T_u(\alpha)^{1-p} = \infty,$$

which, together with (6.9) and the representation (6.11), yields that $\lim_{\log(1+\overline{\lambda}) \searrow C} u(1,0,\overline{\lambda},0) = \infty$. Since, clearly, $u(1,0,\cdot,0)$ is decreasing in $\overline{\lambda}$, this amounts to saying that

$$u(1,0,\overline{\lambda},0) = \infty, \text{ for } \log(1 + \overline{\lambda}) \le C. \qquad \square$$

Remark 6.15. The map $\alpha \to G(\alpha) = \int_\alpha^{\beta_\alpha} \frac{g'_\alpha(x)}{x} dx$ is strictly decreasing in general, not just under the parameters restricted by the hypothesis of Proposition 6.14. The same argument, as the one given in the proof of Proposition 6.14, applies. In particular, this fact can be used to show that the free-boundary problem (2.10), (2.12) has a *unique* solution for all values of the transaction costs, as long as $u < \infty$.

It is, perhaps, interesting to note that the authors are unable to come up with a purely analytic argument for the monotonicity of $G(\alpha)$. The crucial step in the proof of Proposition 6.14 above is to relate the value of $G(\alpha)$ to the original control problem, and then argue by using the natural monotonicity properties of the control problem itself, rather than the analytic description (2.10) only.

References

[Cho12] Jinhyuk Choi, *A shadow-price approach to the problem of optimal investment/consumption with proportional transaction costs and utilities of power type*, Ph.D. thesis, The University of Texas at Austin, 2012.

[CM76] G. M. Constantinides and M. J. P. Magill, *Portfolio selection with transactions costs*, Journal of Economic Theory **13** (1976), 245–263.

[DN90] M. H. A. Davis and A. R. Norman, *Portfolio selection with transaction costs*, Math. Oper. Res. **15** (1990), no. 4, 676–713.

[GK02] Victor Ginsburgh and Michiel Keyzer, *The structure of applied general equilibrium models*, The MIT Press, 2 2002.

[HP11] A. Herczegh and V. Prokaj, *Shadow price in the power utility case*, preprint, http://arxiv.org/abs/1112.4385, 2011.

[JK95] Elyes Jouini and Hédi Kallal, *Martingales and arbitrage in securities markets with transaction costs*, J. Econom. Theory **66** (1995), no. 1, 178–197.

[KMK10] J. Kallsen and J. Muhle-Karbe, *On using shadow prices in portfolio optimization with transaction costs*, Ann. Appl. Probab. **20** (2010), no. 4, 1341–1358.

[KS98] Ioannis Karatzas and Steven E. Shreve, *Methods of mathematical finance*, Applications of Mathematics (New York), vol. 39, Springer-Verlag, New York, 1998.

[LPS98] Damien Lamberton, Huyên Pham, and Martin Schweizer, *Local risk-minimization under transaction costs*, Math. Oper. Res. **23** (1998), no. 3, 585–612.

[Mer69] R. C. Merton, *Lifetime portfolio selection under uncertainty: the continuous-time case*, Rev. Econom. Statist. (1969), 247–257.

[Mer71] _____ , *Optimum consumption and portfolio rules in a continuous-time model*, J. Economic Theory (1971), 373–413.

[Sko61] A. V. Skorohod, *Stochastic equations for diffusion processes with a boundary*, Teor. Verojatnost. i Primenen. **6** (1961), 287–298.

[SS94] S. E. Shreve and H. M. Soner, *Optimal investment and consumption with transaction costs*, Ann. Appl. Probab. **4** (1994), no. 3, 609–692.

[Wal98] Wolfgang Walter, *Ordinary differential equations*, Graduate Texts in Mathematics, vol. 182, Springer-Verlag, New York, 1998, Translated from the sixth German (1996) edition by Russell Thompson, Readings in Mathematics.